Epigenetics of TET2 Loss in Myelodysplastic Syndromes

Epigenetics of *TET2* Loss in Myelodysplastic Syndromes

A dissertation presented

by

Allegra Matheson Lord

to

The Division of Medical Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biological and Biomedical Sciences

Harvard University
Cambridge, Massachusetts

April 2015

Dissertation Advisor: Dr. Benjamin Ebert **Allegra Matheson Lord**

Epigenetics of TET2 Loss in Myelodysplastic Syndromes

Abstract

Myelodysplastic syndromes (MDS) are a class of myeloid malignancy characterized by peripheral blood cytopenias and impaired hematopoietic differentiation. Our understanding of the molecular basis of MDS has improved enormously in recent years due to clinical research efforts to characterize the spectrum of acquired mutations found in patients. This work has revealed that mutations in *TET2* are common lesions in MDS and other myeloid malignancies. TET2 function has only recently been elucidated: TET proteins convert 5'-methylcytosine (mC) first to 5'-hydroxymethylcytosine (hmC), apparently the first step in an active DNA demethylation program that leads to the replacement of 5-mC with unmodified cytosine. My thesis work focuses on a characterization of *TET2* loss on DNA methylation, and on how *TET2* mutations impact patient response to treatment with hypomethylating agents.

We examined DNA methylation in a matched set of *TET2*-WT and -mutant MDS samples, and found that loss of *TET2* results in global hypermethylation. This global increase is due to gains in intragenic methylation, specifically localized to intron-exon boundaries. We then used clonal TF1 cell lines with CRISPR/Cas9-engineered *TET2* mutations to examine global DNA hydroxymethylation. Loss of *TET2* results in a global loss of 5-hmC. By aligning our methylation data with hydroxymethylation data from

TET2-WT cells, we were able to identify direct *TET2* targets. Because changes in mC/hmC with loss of *TET2* appeared to localize to intron-exon boundaries, we investigated the effect of aberrant methylation on mRNA splicing in our TF1 cell system. *TET2* loss resulted in an overall increase in exon skipping, consistent with published data on the effect of methylation on splicing, and hypermethylated regions were enriched for alternate splicing events. These findings suggest that the alterations in hematopoietic differentiation seen in *TET2*-mutant models are due to shifts in the expression of different mRNA isoforms rather than wholesale changes in gene expression. Our data show that loss of TET2 function results in region-specific gains in DNA methylation, and that these alterations affect mRNA splicing by promoting exon skipping. Finally, we have found that presence of *TET2* mutations are positively associated with response to HMA therapy.

Table of Contents

List of Figures and Tables

Abbreviations

5-caC	5-carboxycytosine
5-fC	5-formylcytosine
5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
AML	Acute Myeloid Leukemia
AZA	Azacitidine
BER	Base Excision Repair
CGI	CpG Island
CMML	Chronic Myelomonocytic Leukemia
CMP	Common Myeloid Progenitor
DEC	Decitabine
del(5q)	Chromosome 5q Deletion
DMR	Differentially Methylated Region
DNMT	DNA Methyltransferase
FAB	French-American-British
FACS	Fluorescence-Activated Cell Sorting
GMP	Granulocyte-Monocyte Progenitor
HMA	Hypomethylating Agent
HSC	Hematopoietic Stem Cell
HSPC	Hematopoietic Stem and Progenitor Cell (population)
hUBC	Human Umbilical Cord Blood
IDH1/2	Isocitrate Dehydrogenase 1/2
IPSS	International Prognostic Scoring System
LSK	$Lineage^{lo}$ $Sca\text{-}1^{+}$ $c\text{-}Kit^{+}$
LT-HSC	Long-term HSC
MBD	Methylcytosine Binding Domain
MDS	Myelodysplastic Syndromes
MeDIP	Methyl-DNA Immunoprecipitation
MPN	Myeloproliferative Neoplasm
MS/MS	Tandem Mass Spectrometry
OGT	N-acetylglucosamine (O-GlcNAc) Transferase
Pol II	RNA Polymerase II
RA	Refractory Anemia
RAEB	Refractory Anemia with Excess Blasts
RARS	Refractory Anemia with Ringed Sideroblasts
RRBS	Reduced Representation Bisulfite Sequencing
RRHP	Reduced Representation Hydroxymethylation Profiling
SILAC	Stable Isotope Labeling by Amino Acids in Culture
SNP	Single Nucleotide Polymorphism
ST-HSC	Short-term HSC
TET1/2/3	Ten-Eleven Translocation 1/2/3
VAF	Variant Allele Frequency
WGBS	Whole-Genome Bisulfite Sequencing
WHO	World Health Organization

Acknowledgements

First and foremost, I have to thank Ben Ebert for his support and guidance over the years. Ben is both incredibly smart and incredibly nice, and has an amazing ability to find the positive side to nearly everything—no matter how dejected I might have been about a seemingly failed experiment, talking with Ben always raised my spirits. I always leave his office more excited about science than when I went in. I've really enjoyed being a part of Ben's lab as it has expanded, and while leaving will be bittersweet, I know the lab will continue to be successful due to Ben's leadership and ability to find amazing trainees. Ben also appreciates a good Manhattan, which is an important quality in a mentor.

I have to thank a few of the original Ebert Lab members, who were and remain incredible mentors and friends. Raf Bejar is basically the smartest person I know (next to Ben!), and manages to make it look easy. I'm honored to have been able to collaborate with him, and am looking forward to hearing about his next projects, both scientific and outdoor sports-related. Anu Narla is an astonishingly accomplished scientist, human being, and world traveler. She has been my scientific and personal role model for almost as long as I've known her, ever since we bonded over our mutual dislike of early morning conversation. I owe my title of lab methylcellulose queen to her, as well as my appreciation for prosecco and color-coding as lifestyle choices. Ann Mullally is fond of early morning conversation, but I try not to hold it against her. Ann and I worked closely together in the twilight of the Gilliland lab, and when she joined Ben's lab I followed suit—I couldn't imagine joining a lab in which I wouldn't be tested on my knowledge of early 80's pop music in the tissue culture room, or receive encouragement in the form of impromptu (but always topical) poetry recitations in the animal facility. I'm very lucky to have Ann still close by, as I rely on her scientific expertise, but also her friendship and advice in just about everything. Damien Wilpitz is a dear friend, and a fantastic lab manager—he manages to run an academic lab with an unheard-of level of organization and precision, and we are all completely spoiled by it. On a more personal level, I've valued his constant friendship over the years. Our DnA talks have been a refuge through the very worst times, and our celebrations have made the very best times even better. I won't say that his back rubs are the main reason I joined Ben's lab, but they were probably a factor.

I also have to thank my current lab family—I will miss our late-night dinners, weekend shopping trips, and all-the-time coffee breaks. Rebekka Schneider-Kramann is an amazing post-doc in Ben's lab, and a great friend. I know that I will continue to try to emulate her work ethic in the future, and I have learned an immense amount from her flow cytometry and pathology expertise. Esther Obeng is a fantastic scientist and clinician. I have benefitted scientifically from our conversations about splicing and myeloid disease, but also personally from our mutual motivational talks and shared love of celebrity gossip. Edwin Chen is a very talented post-doc in the Mullally lab, and I've appreciated his dry sense of humor, just Ph.D., and hardly noticeable Canadian accent from day one.

My fellow graduate students have made more bearable or more fun, depending on the day. Peter Miller is basically a genius, but also a generally great person, and I'm very glad to

be able to call him a friend. Rishi Puram's circadian rhythm is the exact opposite of mine, but when we do get to see each other it's always a good time. Dave Raiser has been my supplier of Sour Patch Kids, brotein, and commiseration over the past few years, and I can't wait to see all the ways he'll fix global public health.

When I began graduate school, I was set on joining Gary Gilliland's lab, and while that didn't work out, I was very fortunate to become close with some of the people who made that lab so great, and who have continued to be sources of advice and support. In particular, I'd like to thank Julie Losman, Steve Lane, Steve Sykes, and Mike Kharas for all their help and friendship over the years.

Ben's lab has benefited from amazing technical support over the years, and as with everyone else in the Ebert lab, the techs have been really great people as well as accomplished scientists. I should especially thank Katie Lin, Lisa Chu, and Michelle Chen. All of them have helped with my project in some way, but they have also all been great friends. Many thanks also to Marie McConkey, the Ebert lab staff scientist who has been a part of just about every project in lab at one point or another.

Prior to grad school, I worked in Len Zon's lab with Trista North and Wolfram Goessling. The three of them have remained invaluable mentors through my time in grad school, and I'm very grateful to them (as well as my other Zon lab friends) for their support.

I'd also like to thanks the members of my dissertation advisory committee, Jon Aster, Derrick Rossi, Scott Armstrong, and more recently Alex Meissner, for their advice and discussion over the years. My collaborators made this project possible, and in particular I'd like to thank Kendell Clement for all of his help with bioinformatics. I have learned a ton from him over the years, and I think he's picked up a few things from me as well!

My friends and family have really made my time in grad school possible. My Boston friends—in particular, Teresa, Allison, Tim, Sarah, Bryan, Alex, Diane, and Scott—have all been so supportive through the years, and I'm very lucky to have such a great group of friends here. Life in grad school wouldn't have been the same without our nights out dancing or movie nights with matching cocktails. My friends from my undergrad years have also been wonderful—Molly, Joanna, and Candi all gave me great excuses to spend weekends in New York and hours on the phone. My extended family have made valiant efforts to understand my thesis work, and I'm especially grateful to my grandmother, who is very proud of me and always knew I would be a doctor of some sort one day.

Finally, I have to thank my mother, Tina, and her partner, Joel, above all. They have supported me through everything and been an unending source of love and encouragement, and I am forever grateful.

Chapter 1

Introduction

Myelodysplastic Syndromes

Myelodysplastic syndromes (MDS) are a fairly broad category of myeloid malignancy, or disease of the myeloid lineage of hematopoiesis. In this section, I will discuss hematopoietic differentiation and how it is altered in MDS, clinical characteristics of MDS subtypes and current treatment options, as well as our current understanding of the genetics and molecular mechanisms underlying this disease.

Normal Hematopoiesis

The hematopoietic system is a highly regenerative organ comprised of all mature blood cells and their immature precursors. These cells can be divided into two broad lineages: the lymphoid lineage, which is crucial for adaptive immune response, and the myeloid lineage, which is comprised of cells involved in innate immune response, oxygen delivery, and clotting. All of these cells are derived from a multipotent hematopoietic stem cell (HSC), a rare, usually quiescent cell that resides primarily in the bone marrow (**Figure 1.1**). The HSC is capable of both symmetric and asymmetric division, which allows for both differentiation and maintenance of the HSC pool. HSCs differentiate into multipotent progenitor (MPP) cells capable of reconstituting all hematopoietic lineages but not self-renewal. This cell population gives rise to more lineage-restricted progenitors: the common lymphoid progenitor (CLP) gives rise to cells of the lymphoid lineage but also has some potential to generate cells of the granulocytic/monocytic lineage (Adolfsson et al., 2005), while the common myeloid

progenitor (CMP) gives rise to the granulocyte-monocyte progenitor (GMP) and myelo-erythroid progenitor (MEP).

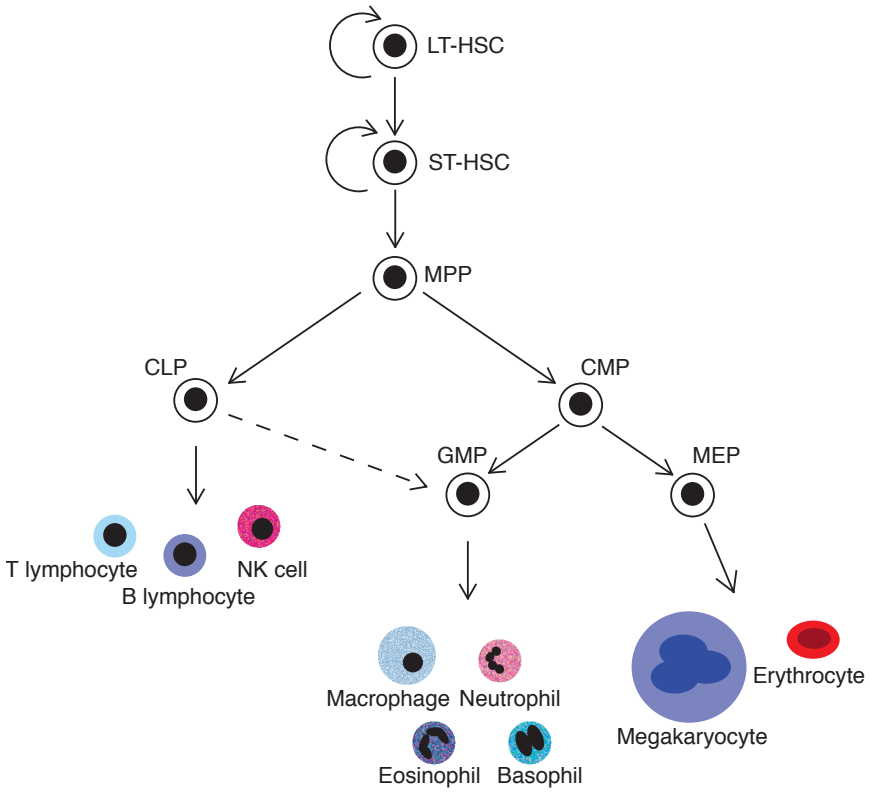


Figure 1.1: The hematopoietic hierarchy.

This simplified diagram shows our current understanding of the hierarchy of hematopoietsis. All hematopoietic cells are derived from a bipotential LT-HSC, which can functionally sustain hematopoiesis over the lifetime of an organism. LT-HSCs give rise to a population of ST-HSCs capable self-renewal as well as generation of all hematopoietic lineages over a short-term period. More differentiated cells are incapable of self-renewal.

Our understanding of hematopoiesis is based on over a century of observation and experimentation. Although the observation that the hematopoietic system is organized as a hierarchy was made in the early 20th century (Maximow, 1909), the first evidence suggesting that a single cell might be capable of regenerating all hematopoietic cells

came in the middle of the 20th century, based on efforts to rescue bone marrow failure following exposure to radiation (Lorenz et al., 1951). Our current understanding of the hematopoietic stem cell and hierarchy is based largely on extensive *in vivo* transplantations in mice, which remain the gold standard model to define hematopoietic stem and progenitor cell potential. These experiments led to phenotypic definitions of populations enriched for hematopoietic stem cell potential, including LSK HSPCs and the current model for enrichment of LT-HSCs using SLAM family markers (Kiel et al., 2005; Yilmaz et al., 2006). Our understanding of human hematopoietic biology was initially limited by technological considerations. The development of immunocompromised mouse models allowing xenotransplantation of human cells has allowed strides to be made in the phenotypic and molecular definitions of human hematopoietic stem and progenitor cells. (Iwasaki and Akashi, 2007)

Phenotypic definitions of hematopoietic stem and progenitor cells have been incredibly useful, but can be limiting in the context of disease and may not fully reflect real biological distinctions between stages of differentiation. Recent work examining transcriptional and epigenetic changes at multiple stages across the hematopoietic lineages has allowed us to gain a deeper understanding of the mechanisms that regulate self-renewal and differentiation. Initial studies in human hematopoietic cells have begun to describe the transcriptional networks regulating hematopoietic differentiation (Novershtern et al., 2011).

Hematopoiesis in MDS

MDS, broadly, are diseases marked by a block in differentiation and a loss of the ability of hematopoietic precursors to form functional mature hematopoietic cells. This results in a hypercellular bone marrow and in peripheral blood cytopenias. Recent work using clinical samples has demonstrated that the HSC is the disease-propagating cell in MDS, and the cell in which initial lesions leading to the development disease are acquired (Woll et al., 2014). These "MDS stem cells" have reduced functional capacity as measured by *in vitro* colony-forming assays, and give rise to dysplastic progenitor cells (Will et al., 2012). Will et al. further demonstrated that the expansion of progenitor cells seen in MDS correlates with disease subtype: low-risk subtypes tended to have an expanded population of the more primitive CMP, while high-risk subtypes showed expansion in the GMP population. MDS progenitors are capable of differentiation to phenotypically mature hematopoietic cells, but these cells are dysfunctional and exhibit increased levels of apoptosis (Albitar et al., 2002).

Clinical Characteristics of MDS

MDS typically (though not exclusively) affects older individuals, with a median age of approximately 70 years (Aul et al., 1998; Rollison et al., 2008). Patients often present with fatigue, anemia, and recurrent infections due to impaired immune function (Hofmann and Koeffler, 2005).

MDS is a highly clinically heterogeneous disorder, and includes multiple subtypes. These are classified based on morphology, and include refractory anemia (RA), refractory anemia with ringed sideroblasts (RARS), and refractory anemia with excess

blasts (RAEB). RA is marked by anemia and a low percentage of blasts (<1% in the peripheral blood, <5% in the bone marrow) and dysplasia in a single myeloid lineage. RARS is similar to RA, but is marked by presence of ringed sideroblasts and dysplasia in the erythroid lineage. RAEB is generally characterized by a higher percentage of blasts (1-19% in the peripheral blood, 5-19% in the bone marrow), as well as dysplasia in one or more myeloid lineages. A further class of myeloid malignancies has predominantly myelodysplastic features with some myeloproliferative characteristics; these are frequently classified as chronic myelomonocytic leukemia (CMML). Finally, MDS can be classified according to the presence of cytogenetic abnormalities. Unlike AML, where chromosomal translocations are more common, deletions and duplications predominate in MDS. The most common of these abnormal karyotypes are deletions within the q arms of chromosome 5 (del[5q]) or 20, loss of chromosome 7, and duplication of chromosome 8. (Adès et al., 2014)

The majority of MDS cases arise with no known cause, though a small fraction are secondary to exposure to chemotherapeutic agents or environmental factors (Adès et al., 2014). Progression from MDS to AML occurs in a substantive fraction of patients, and despite biological differences the two diseases exist on a phenotypic spectrum. The exact mechanism of this transformation remains unclear, though silencing of critical tumor suppressor genes by aberrant DNA methylation has been suggested as a contributing factor (Jiang et al., 2009). Current WHO diagnostic standards differentiate MDS from AML by the percentage of immature cells (blasts) in the bone marrow: MDS is defined as <20% blasts, whereas a blast percentage greater than 20% is defined as AML (Albitar et al., 2002; Swerdllow et al., 2008).

MDS Therapies

Currently the only curative treatment for MDS is allogeneic bone marrow transplantation. However, given the high median age of the MDS patient and potential for complications from this procedure, many patients are instead placed on maintenance therapy to ameliorate symptoms and minimize disease burden (Stone, 2009). Patients with del(5q) respond well to treatment with lenalidomide, a thalidomide derivative that our laboratory has recently shown functions by inducing degradation of Casein Kinase 1a1, which resides on the commonly deleted region of chromosome 5, thereby selectively targeting the del(5q) clone (Sekeres, 2011; Kroenke et al., in revision, 2015). Because alterations in methylation are common in MDS and associated with poor prognosis (Jiang et al., 2009), another option is therapy with the hypomethylating agents (HMAs) azacitidine or decitabine. These drugs induce widespread, non-specific demethylation via DNA methyltransferase inhibition, and improve survival and minimize progression to AML in responsive patients (Garcia-Manero and Fenaux, 2011).

Molecular Basis of MDS

MDS is less well understood at a molecular level than other myeloid malignancies; a quarter of patients have overt chromosomal abnormalities, but until recently the specific genes involved in disease initiation or progression were largely unknown. A notable exception was the subset of MDS patients with del(5q), in which one of the critical deleted genes, RPS14, has been shown to play an important role in erythroid development (Ebert et al., 2008). More recently, a number of large-scale sequencing projects have deepened our understanding of the mutational spectrum in

MDS, and have revealed that the most frequently mutated classes of genes in this disease are involved in mRNA splicing and epigenetic modifications (Table 1.1). Approximately half of all MDS patients have somatic mutations in a spliceosome component, including SF3B1, U2AF1, SRSF2, and ZRSR2, all of which are involved in recognition of the 3' splice site. These mutations tend to be mutually exclusive with other splicing factor mutations, and are typically missense mutations resulting in gains of function. (Pellagatti and Boultwood, 2015)

Table 1.1: Common somatic mutations and their frequencies in MDS (Pellagatti and Boultwood, 2015)

Gene	Frequency in MDS
DNA/chromatin modifiers	
TET2	20-25%
IDH1/2	5%
DNMT3A	10%
ASXL1	10-15%
EZH2	5%
Splicing factors	
SF3B1	20-28%
SRSF2	12-15%
U2AF1	7-9%
ZRSR2	3-11%
Other	
RUNX1	10%
TP53	5-10%
JAK2	5%
NRAS	5%
KRAS	2%
ETV6	2-5%
EVI1	1-2%

Somatic mutations in *Ten-Eleven Translocation 2* (*TET2*), found in 20-25% of patients, are among the most common lesions in MDS, and are also frequently found in other myeloid malignancies including AML. In contrast to mutations in spliceosome components or in members of tyrosine kinase pathways, which are functionally redundant and therefore rarely co-occur in the same patient, *TET2* mutations can be found in combination with nearly every other known mutation in MDS. This implies a novel functional role for *TET2* loss in the development and progression of MDS. The sole exceptions are gain of function mutations in *IDH1/2*, which are rarely found co-mutated with *TET2* mutations. *IDH1/2* mutations are neomorphic, resulting in the production of the oncometabolite 2-hydroxyglutarate (2-HG) which inhibits the function of 2-oxoglutarate-dependent proteins, including TET family members (Figueroa et al., 2010).

TET2 **in MDS**

As discussed above, somatic mutations in *TET2* are a frequent event in MDS. TET2, like other TET family members, converts methylcytosine to hydroxymethylcytosine. The role of loss of this function on MDS development and progression remains unclear. In this section I will discuss the role of Tet2 in hematopoiesis, the function of TET family proteins, and the importance of DNA methylation in hematopoiesis.

TET2 in Hematopoiesis and Hematologic Malignancy

Until recently, of the *TET* family members, only *TET1* had been implicated in myeloid malignancies, as a fusion partner of MLL in some patients with AML (Lorsbach et al., 2003). In the past several years, however, a number of groups have reported *TET2* loss-of-function mutations in primary patient samples of MPN, MDS, CMML, and AML (Abdel-Wahab et al., 2009; Couronné et al., 2010; Delhommeau et al., 2009; Jankowska et al., 2009; Langemeijer et al., 2009; Saint-Martin et al., 2009; Tefferi et al., 2009a; 2009b). In these studies *TET2* has been found to be mutated in CD34+ HSPCs. Studies from our own laboratory have confirmed these findings, and have demonstrated that *TET2* mutations coincide with mutations in a number of other common tumor suppressor mutations in MDS, indicating that *TET2* mutations are likely to affect a non-redundant functional pathway in MDS pathogenesis (Bejar et al., 2011). A recent study reports that acquisition of *TET2* mutations is associated with leukemic transformation in MPN patients who subsequently develop secondary AML (Abdel-Wahab et al., 2010).

Studies in patient samples have been limited by availability and technological constraints. Development of a *Tet2* conditional knockout mouse has been critical to a dissection of TET2 function in hematopoiesis. Multiple models of conditional deletion of *Tet2* in hematopoietic cells have been developed, and show that loss of *Tet2* results in an increase in self-renewal in the HSC compartment, an expansion of the GMP population and corresponding expansion of the granulocyte/monocyte populations, and the development of an MDS- or CMML-like disease (Moran-Crusio et al., 2011; Quivoron et al., 2011). Taken together, these data suggest that loss of TET2 may affect the self-renewal and differentiation pathways of hematopoietic cells.

Function of TET Family Members

TET2 belongs to a family of three 2-oxoglutarate and Fe(II)-dependent dioxygenases, and is expressed in many tissues, including the bone marrow, hematopoietic stem/progenitor cells (HSPCs), and peripheral blood (Bocker et al., 2012; Langemeijer et al., 2009). Tahiliani et al. initially demonstrated that the *TET* family member *TET1* converts 5-methylcytosine to 5-hydroxymethylcytosine (hmC) (Tahiliani et al., 2009). The precise function of 5-hmC is unknown, and will be discussed further below, but it is hypothesized to be a DNA-demethylation signal (Mullighan, 2009). Given our current understanding of the function of DNA methylation, this suggests that *TET* family members may play an important role in the regulation of gene transcription. The oxygenase domains in *TET1* that are predicted to catalyze the transfer of a hydroxyl group are conserved in *TET2* and *TET3,* and recent studies have confirmed conservation of this function in murine *Tet2* (Ito et al., 2010). Although no data have been published on the direct function of human *TET2*, the study of *TET1* and the high level of conservation between functional domains of *TET/Tet* family members support the hypothesis that methylcytosine dioxygenase function is also conserved among all TET family members. Furthermore, myelodysplastic syndrome (MDS) and myeloproliferative neoplasm (MPN) patient samples with *TET2* mutations were shown to have lower levels of global hmC than healthy controls or patients without *TET2* mutations (Jiang et al., 2009).

Two groups have further characterized the function of TET proteins by describing processive reactions in which 5-hmC is converted first to 5-formyl-cytosine (5-fC) and then to 5-carboxyl-cytosine (5-caC) by TET proteins (He et al., 2011; Ito et al., 2010).

These latter two modifications appear to be recognized by the base excision repair (BER) pathway member thymine-DNA glycosylase (TDG), leading to their removal and replacement with unmodified cytosine (He et al., 2011). These findings provide a model for active demethylation of DNA, initiated by TET proteins. Whether or not 5-hmC is a transient mark *in vivo* is controversial: it is unclear whether 5-hmC has a functional role separate from its role as an intermediate in the demethylation process. Interestingly, several methylcytosine binding domain (MBD) proteins may preferentially recognize 5-hmC, while others can bind both 5-mC and 5-hmC, suggesting that 5-hmC may be used to fine-tune the response to DNA modifications (Mellen et al., 2012; Scourzic et al., 2015; Yildirim et al., 2011).

DNA methylation and hematopoiesis

DNA methylation is a means of epigenetic regulation that is increasingly being recognized as critical for the proper specification and differentiation of many tissues, including hematopoietic cells. For the purposes of this thesis, methylated DNA refers specifically to the methylation of the base cytosine, typically found next to a guanine and known as a CpG dinucleotide. Non-CpG methylation, referring to modification of a cytosine adjacent to a base other than guanine, is less well understood (Guo et al., 2013).

Patterns of DNA methylation are set up during embryogenesis by DNA methyltransferase (DNMT) 3a and DNMT3b, which are capable of methylating cytosine *de novo* (Hsieh, 2000). Methylation is lost as DNA replicates at each cell division, and must be maintained by DNMT1, which has higher affinity for hemimethylated DNA (Bashtrykov et al., 2012). DNA methylation must therefore be actively maintained, and

can be lost passively through a decrease in DNMT1 activity. The existence of an active DNA demethylation program in mammals has been postulated for many years, but most putative mammalian demethylase mechanisms have either been disproven or never replicated (Ooi and Bestor, 2008). More recent work implicating the TET family proteins in DNA demethylation will be discussed below.

The regulation of DNA methylation is critical to understand because of the importance of this modification in the regulation of transcription and therefore cell state. DNA methylation is typically found concentrated in CpG rich regions called CpG islands (CGI) and at gene promoters, though more recent work has explored the importance of methylation outside these areas, in regions called CGI shores (within 2kb of a CGI) or shelves (2kb upstream or downstream of CGI shores) (Álvarez-Errico et al., 2015), as well as in gene bodies. Nevertheless, CGI and promoter methylation remains the best understood in terms of function: heavily methylated CGI and promoters are associated with repression of transcription of the associated gene. Transcriptional repression through methylation is achieved in two ways. First, methylcytosine can directly inhibit binding of transcription factors and other DNA-binding proteins. Second, methylcytosine can be bound by various methyl-binding domain (MBD) proteins, which in turn act as inhibitors of transcription. (Schübeler, 2015)

Methylcytosine-mediated gene silencing is important during development, but also in the differentiation of mature cells from precursors. Methylation dynamics during hematopoiesis have been well studied in the mouse, and reveal that methylation patterns of HSCs are more closely related to myeloid lineage cells than to lymphoid lineage cells, implying that HSCs have some inherent myeloid bias (Beerman et al., 2013). Methylation

overall increases as HSCs differentiate, with homeobox genes in particular becoming repressed through increased methylation (Bock et al., 2012). In myeloid lineage cells, however, changes in methylation are more dynamic. Although DNA methylation increases initially in the transition from HSC to CMP, terminally differentiated myeloid cells are typically hypomethylated relative to HSCs (Álvarez-Errico et al., 2015). Loss of maintenance methylation in HSCs in a conditional *Dnmt1* knockout mouse diminishes self-renewal and engraftment capacity as well as specifically myeloid lineage differentiation, highlighting the functional importance of DNA methylation in hematopoiesis (Trowbridge et al., 2009)

Summary/Rationale

Loss of *TET2* is a frequent event in myeloid malignancies, including MDS, and evidence from murine models suggests that Tet2 deficiency affects HSC function and functional capacity. Lesions in *TET1* or *TET3* are comparatively rare in hematopoietic malignancies (Abdel-Wahab et al., 2009; Atak et al., 2012; Lorsbach et al., 2003). This may be due to tissue specificity of the TET family members, but also points to a unique role for TET2 in hematopoiesis and myelopoiesis.

Given the known function of TET family members as methylcytosine dioxygenases, a logical hypothesis follows that loss of TET2 function would lead to global and/or site-specific DNA hypermethylation. Results from initial studies on the effect of *TET2* loss on DNA methylation have been mixed, however, possibly due to methodology or confounding factors affecting methylation status (Figueroa et al., 2010;

Ko et al., 2010). Furthermore, no group has examined changes in hydroxymethylation relative to changes in methylation, or the functional effects these epigenetic changes may have in hematopoietic cells. For these reasons, I aimed to carefully assess whether and how DNA methylation changes with loss of *TET2* in primary MDS samples, as well as in genetically engineered model systems, and to link any changes in methylation to hydroxymethylation and other functional changes. Finally, given our hypothesis regarding the effect of *TET2*-loss on DNA methylation, we sought to determine whether the presence of *TET2* mutations might affect response to HMAs.

Chapter 2

Materials & Methods

MDS patient samples

Bone marrow aspirate mononuclear cells from MDS patients were obtained as described (Bejar et al., 2011). Patients provided written informed consent according to protocols developed by the institutional review board at the institution collecting the samples. Genomic DNA was stored at the Biological Samples Platform at the Broad Institute in Cambridge, MA.

CD34+ Cell Experiments

Isolation

Mononuclear cells were isolated from human umbilical cord blood (National Cord Blood Program, NYBC)) by density gradient centrifugation with Ficoll-Paque. CD34+ stem/progenitor cells were isolated from total mononuclear cells using anti-CD34 antibody and the autoMACS Pro Separator (Miltenyi).

Culture Conditions

hUBC-derived CD34+ cells were cultured in StemSpan^{TM} SFEM (Stemcell Technologies) with Penicillin/Streptomycin (100 U/mL), glutamine (2 mM), lipids (40 μg/mL), TPO (50 ng/mL), SCF (25 ng/mL), and FLT-3 (40 ng/mL) (FST media).

shRNA Knockdown

Lentivirus was produced according to standard protocols from pLKO containing shRNA against *TET2* (TRC 122172: GCGTTTATCCAGAATTAGCAA; TRC 144344: CCTTATAGTCAGACCATGAAA) or nontargeting control shRNA, and concentrated via ultracentrifugation. Thawed hUBC-derived CD34+ cells were plated in 48-well format 24 hours post-thaw in 100 μL FST media with 2x cytokines and polybrene (8

μg/mL), and mixed with 100 μL concentrated virus. Plates were spun at 1500RPM for 60 minutes at 23°C, then incubated at 37°C for 1 hour, after which 200 μL FST media was added and cells were cultured overnight. Antibiotic selection was performed for 3 days with puromycin (2 μg/mL).

TF1 cells

TF1 erythroleukemia cells were grown in RPMI supplemented with 10% FBS, 1% Penicillin/Streptomycin, and GM-CSF (2ng/mL). To generate *TET2*-mutant TF1 cells, we transduced cultures with a single lentiviral vector containing a puromycin selection marker, Cas9 and a guide RNA targeting either *TET2* exon 3 (Supplemental Figure 4) or GFP (control) as described (Sanjana et al., 2014). Cells were placed under puromycin selection 24 hours after infection, and grown under selection for one week. Single cells were then sorted into 96-well plates and single cell clones were expanded for three weeks. Genomic DNA from clonal lines was extracted and sequenced using primers surrounding the gRNA target sequence.

Methylation profiling and analysis

RRBS was performed as described (Gu et al., 2010; Meissner, 2005). Global methylation was determined by averaging relative methylation across 1kb tiles for all well-covered (≥10 reads) regions. Regional methylation was determined in the same way, with the following changes: promoters were defined as the 2kb region centered at the transcription start sites of Refseq genes, CGIs were defined using a biologically-verified set (Illingworth et al., 2010), enhancers were defined as H3K27ac peaks in CD34+ cells

from Roadmap samples (GSM 772885, GSM772870, GSM772894) or Vista dataset, intron-exon regions were defined as the 1kb region centered at non-promoter intron-exon junctions of RefSeq genes. WGBS was performed as described (Gifford et al., 2013).

RNA Sequencing

Total RNA was extracted from 4 MDS samples stored in GITC buffer in an RNAse-free environment using standard protocols, and the product of RNA extraction was cleaned up using the PrepEase RNA Spin kit (Affymetrix). Total RNA from TF1 cell lines was extracted using the PrepEase RNA Spin Kit. RNA sequencing was performed on all samples using the Illumina TruSeq platform to a depth of ≥50M strand-specific paired reads. Reads were aligned to the reference genome using the Picard pipeline. Whole transcriptome analysis was performed using Cufflinks, and quantification of alternate splicing events was performed using JuncBASE.

MeDIP PCR

Genomic DNA from TF1 cell lines was extracted (Qiagen Blood Mini) and digested with MseI according to standard protocols. MeDIP was performed in triplicate for each cell line using the MethylCollector Ultra kit (Active Motif) according to the manufacturer's protocols. Primers for individual DMRs were designed to fall within 5' and 3' MseI cut sites most proximal to each DMR. Quantitative PCR was performed using the products of MeDIP and DMR-specific primers, and products measured with SYBR-Green. Each reaction was performed in triplicate, and was normalized to input DNA to obtain a final value for total methylation at each DMR.

5-hmC detection

Dotblots

Genomic DNA from TF1 cell lines (*TET2*-wildtype or –deficient) was extracted (Qiagen Blood Mini) and digested with MseI according to standard protocols. Fragmented DNA (400ng/sample) was denatured (NaOH final concentration 0.1M, 95C, 5') and then cooled on ice with the addition of ammonium acetate (0.66M final concentration). The dotblot apparatus (Bio-Rad 170-6545) was assembled and samples were loaded and transferred to a rehydrated nitrocellulose membrane according to manufacturer instructions. Samples were crosslinked to the membrane (UV Stratalinker, Stratagene), which was blocked overnight (10% milk, 1% BSA in PBS + 0.1% Tween). The membrane was incubated with primary antibody to 5-hmC (Diagenode C15220001, 1:500) and an HRP conjugated secondary antibody (Santa Cruz Biotech sc-2032) with standard washes.

hmC Sequencing (RRHP) Library Preparation

Genomic DNA was extracted from TF1 cell lines as described above. Genomic DNA was fragmented overnight at 37°C with a hydroxymethyl-insensitive enzyme, MspI, and purified using the DNA Clean and Concentrator kit (Zymo Research). Modified Illumina TruSeq P5 and P7 adapters containing 5'-CG overhangs were ligated onto the digested DNA using T4 DNA ligase (2 hours at 16°C). Libraries were then strand extended at 72°C with Taq DNA Polymerase. The adapters were designed to regenerate the 5'-CCGG site at the P5 junction while the P7 adapter generates a 5'-TCGG junction, making it insensitive to MspI digestion. Adapterized libraries were treated with β-

glucosyltransferase to label 5-hmC modifications and purified using the DNA Clean and Concentrator kit. The glucosylated libraries were then subjected to an overnight MspI digestion at 37°C, cutting any fragments not containing a glucosyl-5hmC site at the P5 CCGG junction. After incubation, the libraries were size-selected from 100bp to 500bp and purified using the ZymoClean Gel DNA Recovery Kit (Zymo Research). The fragments were amplified using OneTaq 2X Master Mix (NEB), and the PCR conditions include an initial denaturation of 94°C for 30 sec followed by 12 cycles of 94°C for 30 sec, 58°C for 30 sec, and 68°C for 1 min. Fragments containing 5-hmC were positively selected during PCR amplification with adapter-specific indexing primers whereas fragments lacking glucosylated-5-hmC at the P5 junction were cleaved and, therefore, not amplified by PCR. Amplified libraries were purified using the DNA Clean and Concentrator kit, and multiplexed using equal volume of the libraries. All adapters and primers used were synthesized by Integrated DNA Technologies.

Bioinformatics processing and statistical analyses

Sequence reads from RRHP libraries were first processed to trim off the low quality bases and the P7CG adapter at the 3' end of the reads. Reads were then aligned to the reference genome using the Bowtie default parameters and --best. Aligned reads with the MspI tag (CCGG) were counted. The correlation analysis between different RRHP libraries was performed by comparing the presence of the tagged reads at each profiled MspI site, and Pearson's coefficient was calculated accordingly.

Patient Samples and Hypomethylating Agent Response Assessment

A total of 213 MDS patients treated with azacitidine or decitabine were included in this study. Samples were obtained from patients treated at the Dana-Farber Cancer Institute (2003-2010, n = 42), the MD Anderson Cancer Center (2003-2010, n = 104), and as part of the DACO-020 (ADOPT) clinical trial of decitabine (2005-2006, n = 67). All samples were collected with patient consent under IRB-approved protocols in accordance with the Declaration of Helsinki. Response to treatment was assessed using International Working Group (IWG) response criteria revised in 2006. Patients with either a complete response (CR), partial response (PR), or hematologic improvement (HI) were considered as 'responders' (R, n = 100, 47%). whereas patients described as having 'no response', 'stable disease', 'progressive disease', 'death' before response assessment, or 'not evaluable' were considered 'non-responders' (NR, n = 113, 53%).

Sample Processing, DNA Sequencing, and Mutation Analysis

DNA was extracted from bone marrow mononuclear cells or peripheral blood samples collected prior to treatment (median 18 days, range 9 to 119). Whole genome amplification of DNA for each sample was performed using the REPLI-g kit from Qiagen. A genotype fingerprint of 22 common single nucleotide polymorphisms (SNP) for each sample was generated by MALDI-TOF genotyping (Sequenom). Target regions of 40 genes (**Supplemental Table 4.1**) and genotype fingerprint regions were enriched using Agilent SureSelect hybrid capture system according to manufacturer instructions. Barcoded samples were pooled in equimolar amounts and subjected to 100 nucleotide paired-end sequencing on an Illumina Hi Seq 2000. Sequence reads were aligned to the

human genome (Build 37) using the Burroughs-Wheeler algorithm (Li and Durbin, 2010). The Genome Analysis Toolkit was used to clean and locally realign reads prior to calling missense and insertion/deletion variants using MuTect (Cibulskis et al., 2013; DePristo et al., 2011). Sample identity was confirmed by matching fingerprint genotype calls. Synonymous variants, non-coding variants more than 6 bases from splice junctions, or variants present in databases of "normal" genomes (dbSNP 132 or NHLBI Exome Sequencing Project) at a population frequency of 1% or more were discarded. Remaining variants were considered candidate somatic mutations.

Competitive Murine Bone Marrow Transplants

Age-matched $Tet2^{fl/fl}$; $Mx\text{-}Cre^{+}$ and $Tet2^{+/+}$; $Mx\text{-}Cre^{+}$ donor animals (CD45.2) were treated with pIpC (15μg/g IP) for three non-consecutive days to induce excision of exon 3 of *Tet2* (Moran-Crusio et al., 2011). Donor bone marrow was harvested two weeks post-pIpC and mixed in a 1:2 ratio with bone marrow harvested from 45.1 WT donors (B6.SJL-Ptprca Pepcb/BoyJ; Jackson), and was then transplanted into 45.1 recipients for a total of 1 million cells/recipient. Peripheral blood engraftment was assessed by FACS at two weeks post-transplant, at which point recipient mice were divided into treatment groups (n=7 per group) and treated daily with either 5-azacitidine (AZA; 2.5mg/kg IP; Santa Cruz Biotechnology) or vehicle control on the following schedule: two weeks on, two weeks off. Peripheral blood chimerism and CBC were assessed following each round of treatment.

Statistical Methods

Categorical variables were compared using a Fisher exact test or chi-square test as appropriate, while continuous variables were compared using a Wilcoxon rank-sum test. A Cochran-Mantel-Haenszel (CMH) test was used to test for differences in response rate by mutational status controlling for treatment. Unadjusted and adjusted logistic regression models were used to predict response to therapy. Models were adjusted for covariates including age (≥70 yrs. vs. <70 yrs.), sex, International Prognostic Scoring System (IPSS) risk group (low/intermediate 1 vs. intermediate 2/high) and treatment (azacitidine vs. decitabine alone vs. decitabine ± other). The odds ratio (OR) and 95% confidence intervals (CI) were estimated for the risk group (mutated) and compared to the reference group (wild-type). The Hosmer and Lemeshow goodness-of-fit test was used to assess model fit of logistic regression models. Overall survival was calculated from the time of treatment to the time of death from any cause or was censored at the date last known alive and was compared using a log rank test. Unadjusted and adjusted univariate Cox models were also constructed using with the same covariates. For the competitive murine experiments, the percent 45.2 chimerism was calculated for each time point. Error bars indicate standard error of the mean (SEM) for each group, and p-values for each time point were calculated using a two-sample t-test. All p-values reported are two-sided and considered significant at the 0.05 level. No adjustments were made for multiple hypothesis testing.

Chapter 3: Results – Part 1

TET2 loss in myelodysplastic syndrome results in intragenic hypermethylation and increased exon skipping

Contributions

I selected patient samples for methylation profiling, and, in collaboration with others (see below), developed approaches for analysis of all genomic/transcriptomic data. I generated the *TET2*-null TF1 cell system used for hydroxymethylation profiling and RNA-sequencing, and performed all validation experiments for genomic/transcriptomic studies. I performed all *in vitro* hUBC-derived CD34+ cell and murine experiments.

MDS patient samples used in this work were first characterized by Rafael Bejar (Bejar et al., 2011), who also advised on sample selection. MDS samples were obtained from Azra Raza's laboratory. All methylation profiling experiments were done in collaboration with members of Alex Meissner's laboratory, notably Hongcang Gu, who performed the bisulfite conversion and sequencing, and Kendell Clement, with whom I worked closely to develop approaches to analyze the data generated and who performed all computational work. 5-hmC capture and sequencing for hydroxymethylation profiling (RRHP) was done at Zymo Research. RNA-sequencing was done through the Genomics Platform at the Broad Institute, and analysis of results was done in collaboration with Michael Burger and Aviad Tsherniak. Angela Brooks (Matthew Meyerson laboratory) provided invaluable advice about use of the JuncBASE analysis platform for alternate splicing events. Rebekka Schneider-Kramann, Ann Mullally, and Michelle Chen provided technical support for murine experiments.

Note: We are currently preparing a manuscript for publication based on work presented in this chapter.

Summary

The methylcytosine dioxygenase TET2 is among the most commonly mutated genes in myelodysplastic syndromes (MDS), but the genomic consequences of TET2 mutation in myeloid transformation remain unclear. The effect of *TET2* loss on DNA methylation in human patients is controversial. We examined the DNA methylation status of a well-characterized cohort of MDS patients (n=66) and found increased global methylation in *TET2* mutant samples versus *TET2*-wildtype samples. We found no difference in methylation between groups at promoters or within enhancers, but *TET2*-mutant samples had increased intragenic methylation, particularly at intron-exon boundaries. Furthermore, regions hypermethylated in *TET2*-deficient samples are enriched for hydroxymethylcytosine in *TET2*-WT cells. Analysis of RNA sequencing revealed an increase in exon-skipping events in *TET2*-null cells compared with WT, specifically involving regions of aberrant hypermethylation in *TET2* mutant patients. Our data suggest that loss of *TET2* results in aberrant hypermethylation at intron-exon boundaries, which in turn alters mRNA splicing.

Introduction

TET family members convert 5-methylcytosine to 5-hydroxymethylcytosine (Tahiliani et al., 2009), which is then thought to be further modified and replaced with unmodified cytosine via the base excision repair (BER) pathway (He et al., 2011; Ito et al., 2011).

TET2 and family members are therefore important regulators of DNA methylation. High levels of DNA methylation in CpG islands near promoters have long been associated with the repression of transcription (Klose and Bird, 2006). DNA methylation at enhancers has also been shown to repress transcription, and is dysregulated in cancer (Aran et al., 2013). Somatic mutations in *TET2* leading to loss of methylcytosine dioxygenase function (Ko et al., 2010) are extremely common in a wide variety of myeloid malignancies, including myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML), and are associated with acquired somatic mutations in spliceosome components such as *ZRSR2* and *SRSF2* (Haferlach et al., 2013; Hanssens et al., 2014; Papaemanuil et al., 2013).

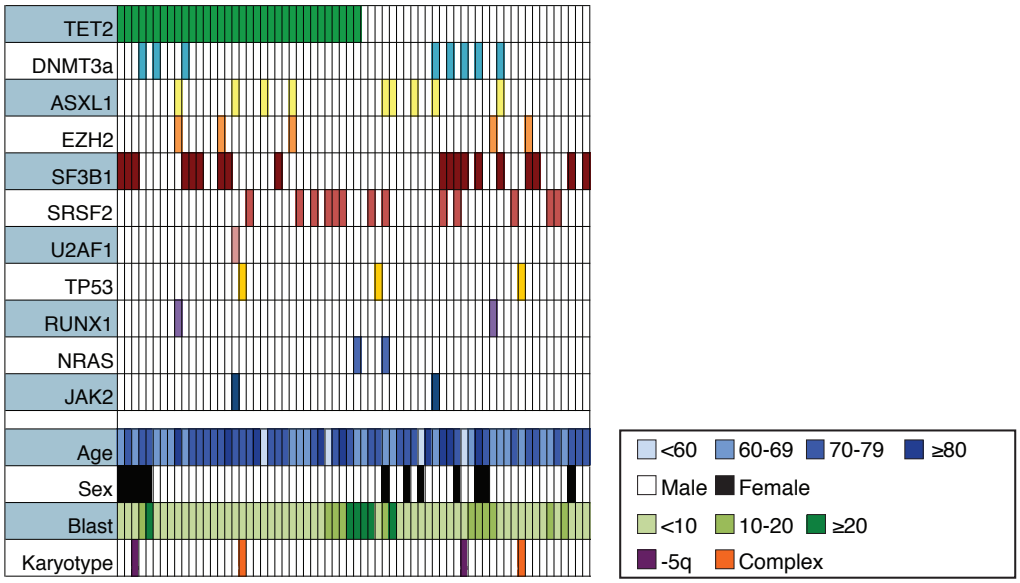
Previous studies examining the effect of *TET2* loss on DNA methylation in myeloid malignancies have reached conflicting conclusions. Ko et al. investigated changes in methylation status with *TET2* loss in a set of patients with varying myeloid malignancies (MDS, MPN, primary and secondary AML) and found overall hypomethylation of *TET2*-mutant samples when compared to *TET2*-WT samples (Ko et al., 2010). This study used the Illumina Infinium array to assess changes in methylation, which is hampered by limited genomic coverage (Bock et al., 2010). Figueroa et al. used the HELP assay to assess methylation and found that AML patients with *TET2* mutations have increased methylation when compared to normal bone marrow (Figueroa et al., 2010). The discrepancy in findings between these two studies may be due to the use of different assays to detect DNA methylation, the presence of confounding somatic mutations with the potential to affect DNA methylation, or the choice of controls.

Results

We sought to characterize the effects of somatic *TET2* mutation on DNA methylation in a well-annotated set of samples from a single disease, MDS, by reduced-representation bisulfite sequencing (RRBS). We analyzed genomic DNA isolated from bone marrow aspirates from 66 MDS cases matched for *TET2* mutation status, somatic mutations in other genes, MDS subtype, blast counts, sex, and age (**Table 3.1, Supplemental Table 3.1**). Genome methylation analyses showed that MDS samples were highly methylated regardless of genotype (**Supplemental Figure 3.1a**), but we observed a shift towards increased global methylation in the *TET2*-mutant samples versus WT (**Supplemental Figure 3.1b**).

Table 3.1: Somatic Mutations in and Clinical Characteristics of MDS Samples.

Columns represent individual patients, while rows describe somatic mutations or patient characteristics. -5q: del(5q); Complex: ≥3 chromosomal aberrations.



In the bone marrow of MDS patients, the percentage of cells bearing *TET2* mutations vary, and these mutations can be either heterozygous or homozygous. To minimize the dilutive effects of cells without *TET2* mutations, we performed a focused analysis of the 17 samples with a *TET2* mutant allele fraction of >40% and matched samples without *TET2* mutations (Bejar et al., 2011) (**Supplemental Table 3.2a, Supplemental Figure 3.2**). Consistent with the observed effects of TET2 mutations, we found that the samples with higher *TET2* mutant allele fraction had a greater increase in methylation at regions of high overall methylation, providing further evidence that the observed shift in methylation is due to *TET2* loss (**Supplemental Figure 3.1c, d**). We binned these results by relative methylation level to compare the number of regions with low, intermediate, and high methylation within the *TET2* mutant and WT groups, and found that *TET2*-mutant samples had significantly increased frequency of regions with high levels of methylation (0.8-1.0) offset by a concomitant decrease in the frequency of regions with intermediate methylation (0.6-0.8) (**Figure 3.1a**). These results indicate that loss of *TET2* leads to increased global methylation in a dose-dependent manner.

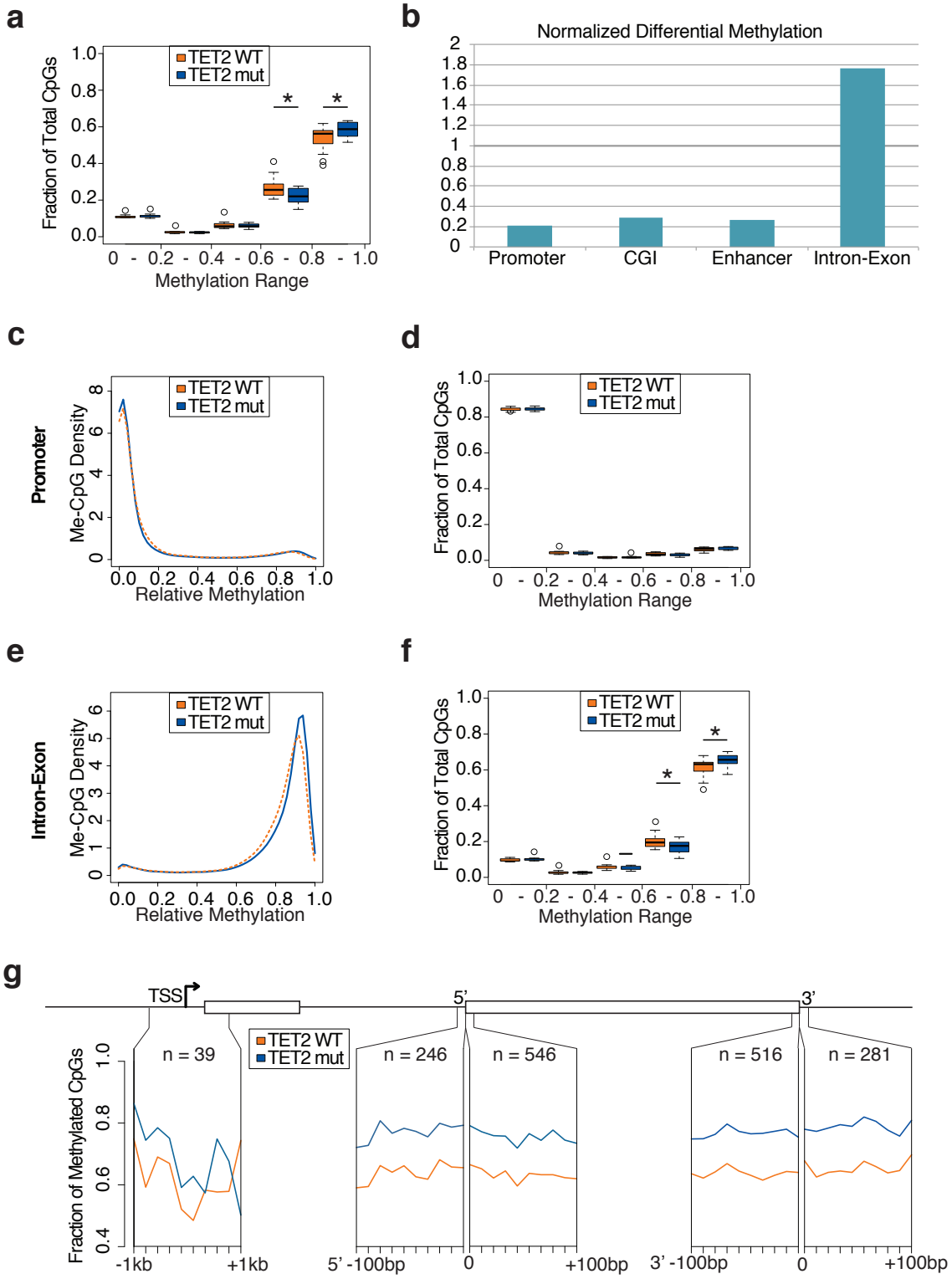
We next sought to determine the genomic regions most affected by *TET2* mutation. To control for differences in region size we normalized differential methylation in each region to CpG content (**Figure 3.1b**). While the majority of total CpGs were located in promoters, CpG islands, and enhancers, these regions contained only a small fraction of differentially methylated CpGs, and differential methylation in these regions was evenly split between the *TET2*-mutant and *TET2*-WT samples (**Supplemental Table 3.3**). Although promoters are enriched for 5-hmC, and Tet family binding in murine ES cells (Wu et al., 2011), we found no significant difference in methylation between groups

at promoters, which had low levels of methylation overall (**Figure 3.1c, d**). Methylation at enhancers has been implicated in regulation of transcriptional control (Aran et al., 2013), but we saw no significant differences between *TET2* mutant and WT groups in the overlap of our methylation data and two separate enhancer datasets (**Supplemental Table 3.3**). In contrast, non-promoter intron-exon boundaries showed marked enrichment for differential methylation despite containing a relatively low fraction of total CpGs (**Figure 3.1e, f**).

Figure 3.1: TET2 mutant MDS samples are hypermethylated vs WT.

Well-covered CpGs were binned according to level of relative methylation and statistics were performed between TET2 mutant and TET2 WT groups for each bin (e; paired t-test, * indicates $p<0.05$) (a); Differential methylation was normalized to region size for the regions indicated, and was calculated as the fraction of differentially methylated CpGs in a given region relative to the fraction of total CpGs in that region (b); Promoter methylation (c,d) and intron-exon methylation (e, f) shown as comparison of average methylation between TET2 mutant and TET2 WT groups (c, e) and relative methylation of each group as a function of density (d, f); Differential methylation across the promoter region and the intron-exon region was averaged and plotted as a function of distance from the TSS and the 5' and 3' exon boundary, respectively.

Figure 3.1, continued



One limitation of these analyses is that the MDS samples available to us were genomic DNA from whole bone marrow aspirates containing a heterogeneous mixture of cells. We therefore sought to confirm our previous findings in genetic models of *TET2/Tet2* loss. We generated *TET2*-deficient hUBC-derived CD34+ HSPCs using an shRNA knockdown approach, and isolated LSK cells from pI:pC-treated *Mx-Cre*; $Tet2^{fl/fl}$ mice. The methylation status of genomic DNA from these cells and corresponding wildtype cells was analyzed by RRBS as before. Global methylation and intron-exon methylation were both increased in the *TET2/Tet2*-deficient cells versus wildtype, while promoter methylation was unchanged (**Supplemental Figure 3.3**). While these results did not reach statistical significance due to small sample size, the trends seen in methylation changes in genetic models of *TET2* loss reinforce our previous findings in MDS samples.

The finding that intron-exon boundaries contained a high relative proportion of differential methylation in our samples is consistent with reports that intron-exon boundaries are enriched for 5-hmC in neuronal tissues (Khare et al., 2012). We examined the precise location of differential methylation relative to the intron-exon boundary by plotting the relative methylation level of well-covered differentially methylated CpGs within 100bp of an intron-exon boundary relative to distance the intron exon boundary. We saw no bias in differential methylation towards 5' versus 3' boundaries, but observed a consistent gain in methylation in the *TET2* mutant group across this region, with a greater absolute number of DMRs in exons versus introns proximal to intron-exon junctions (**Figure 3.1g**).

Methylation profiling by RRBS is designed to analyze CpG-rich regions, and although we obtained good coverage relative to RRBS standards, only 17% of intron-exon junctions in the human genome were analyzed in our data (**Supplemental Figure 3.4**). To examine TET2-dependent methylation at all intron-exon junctions, we performed whole-genome bisulfite sequencing (WGBS) on two MDS patient samples with *TET2* mutations and two without *TET2* mutations (**Supplemental Table 3.4**). WGBS results confirmed the RRBS results in all regions, demonstrating a significant gain of methylation in the *TET2*-mutant samples (**Supplemental Figure 3.5a**), but no differences in methylation at promoters (**Supplemental Figure 3.5b**). At intron-exon boundaries, we again observed an overall significant increase in methylation in *TET2*-mutant samples (**Figure 3.2a, b**). WGBS results for specific regions were concordant with results obtained previously by RRBS, and showed expanded regions of differential methylation concordant with hmC content in *TET2*-wildtype cells (**Figure 3.2d-g**).

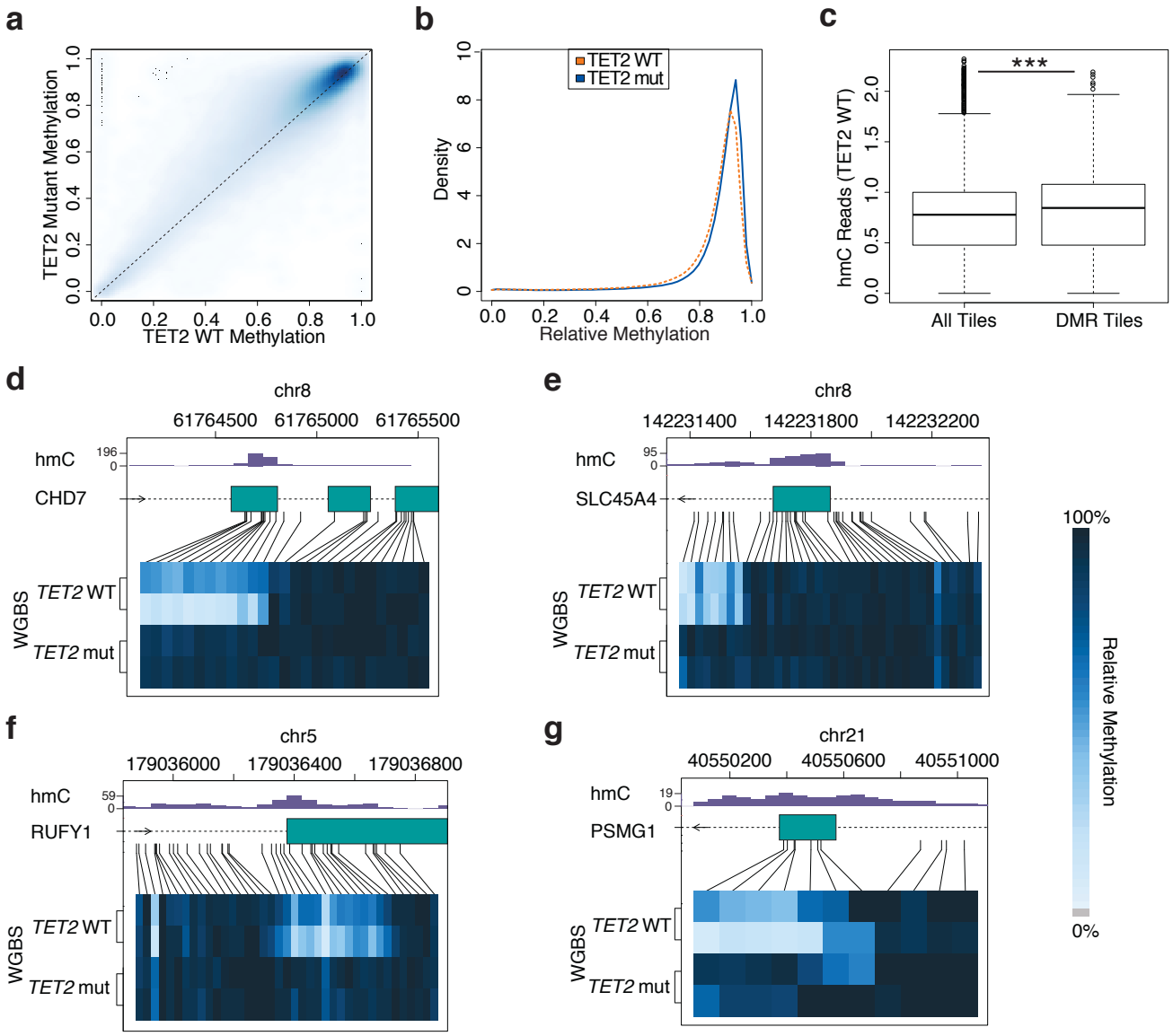


Figure 3.2: Intron-exon hypermethylated regions in *TET2*-mutant samples are enriched for hydroxymethylcytosine in *TET2*-wildtype samples.
Comparison of intron-exon WGBS results between *TET2*-wildtype and *TET2*-mutant groups, where darker blue indicates increased density (a), and WGBS intron-exon methylation in *TET2*-wildtype and *TET2*-mutant groups as a function of density (b). hmC reads by RRHS in *TET2*-wildtype TF1 cells were compared to hypermethylated intron-exon regions ("DMR tiles") and all intron exon regions ("All Tiles") (c; two sample t-test, *** indicates $p \leq 0.0001$). WGBS methylation data and RRHS hmC data (in *TET2*-wildtype) are shown for selected regions (d-g).

To investigate changes in methylation seen with loss of *TET2* in MDS patient samples, we generated two clonal *TET2*-null TF1 cell lines with targeted, homozygous disruption of the *TET2* gene using CRISPR-Cas9 (**Supplemental Figure 3.6**). Both clones had reduced TET2 function as assessed by overall 5-hmC content (**Supplemental Figure 3.6e**). To identify specific sites of altered 5-hmC content, we performed 5-hmC sequencing using an 5-hmC-pulldown approach (RRHS, Zymo) on DNA from these cell lines and a *TET2*-WT TF1 cell line transduced with non-targeting gRNA. Consistent with an overall loss of 5-hmC in *TET2*-null cells, we observed a tenfold increase in unique sites captured coupled with an average tenfold increase in reads per region in the *TET2*-WT samples compared to *TET2*-null samples. Interestingly, approximately half of the sites captured with significant coverage in the *TET2*-null samples mapped to the mitochondrial chromosome, which contains both 5-mC and 5-hmC (Bellizzi et al., 2013; Manev et al., 2012).

To ensure that methylation changes were consistent between MDS samples and the TF1 lines, we validated a subset of the most differentially methylated DMRs from MDS samples in TF1 lines by methyl-DNA-immunoprecipitation (MeDIP) followed by qPCR (**Supplemental Figure 3.7**). To identify putative direct *TET2* target sites, we then examined the overlap of RRHS and WGBS sequencing data. Interestingly, although a number of hmC reads corresponded to regions of hypermethylation, the majority of overlapping sites occurred in regions identified as non-differentially methylated between *TET2*-mutant and -WT samples by bisulfite sequencing (**Supplemental Figure 3.8)**. This result may be an artifact of the inability to discriminate between 5-hmC and 5-mC by

bisulfite sequencing, which would point to the possibility that 5-hmC is a durable modification *in vivo*, or it may be the result of intact TET1 activity in these cells. We compared the overlap of sites with significant coverage by RRHS with all regions covered by WGBS and with DMRs, and saw a significant increase in 5-hmC in *TET2*-WT cells in intron-exon regions hypermethylated in *TET2*-mutant samples (**Figure 3.2c**).

To investigate the link between intragenic methylation and alterations in transcription, we performed RNA sequencing on *TET2*-null and -WT TF1 cell lines. Whole-transcriptome analysis revealed few significant changes in gene expression between *TET2*-WT and -null cells (**Supplemental Figure 3.9**). We next examined alternative splicing events using the JuncBASE analysis pipeline (Brooks et al., 2011). We first examined the frequency of categories of alternative splicing events, irrespective of directionality, in *TET2*-null and -WT TF1 cells. The frequency of events in *TET2*-null cells was similar to baseline (WT vs WT) for all classes except cassette exon inclusion (**Figure 3.3a**). We examined directionality of alternative splicing events and found a significant changes in inclusion reads in the cassette, coordinate cassette, and mutually exclusive exon categories, indicating an increase in exon-skipping in *TET2*-null cells (**Figure 3.3b**).

We then specifically examined the alternative splicing events in *TET2* target regions. We generated a test gene set using non-promoter intron-exon DMRs from WGBS data also enriched for hmC (≥5 reads) in *TET2*-wildtype cells by RRHS. A control gene set composed of highly methylated (normalized methylation ≥ 0.8) but non-differentially methylated non-promoter intron-exon regions (non-DMR) was generated as

a comparator. We then examined splicing events (FPKM $\geq \log_2(-3)$, $q \leq 0.05$) in each gene set, and found a positive enrichment for alternate splicing in the DMR gene set versus the non-DMR gene set (OR: 1.463642, p = 1.255e-05). Genes containing regions hypermethylated in *TET2*-mutant samples and hydroxymethylated in *TET2*-wildtype cells that are also involved in an exon-skipping event are shown in Figure 3c, and are enriched for hematopoietic regulators and tumor suppressors.

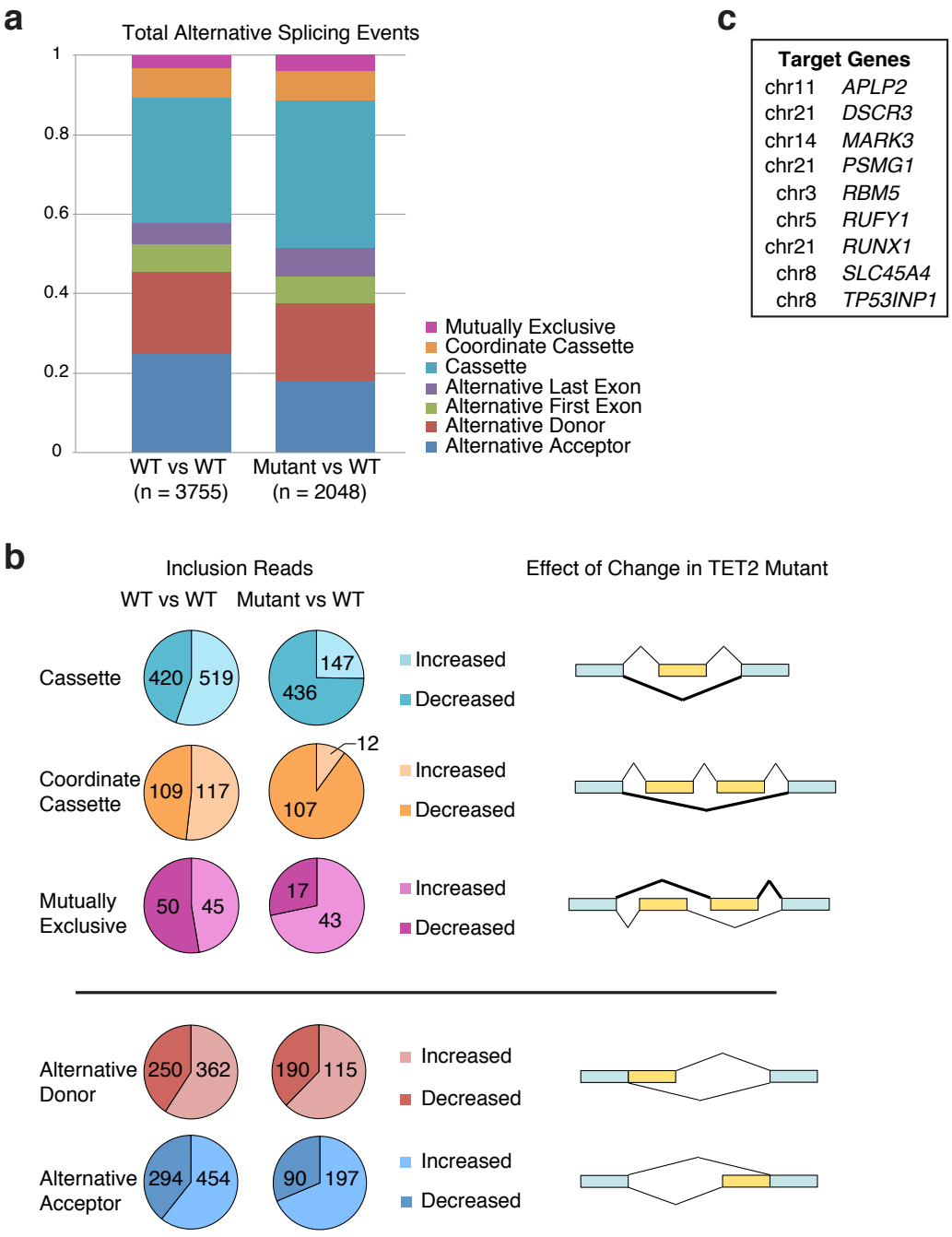


Figure 3.3: Increased exon skipping in *TET2*-mutant samples.
Baseline alternate splicing was examined by comparing alternate splicing events in *TET2*-wildtype replicates (WT vs WT change ≥ 10%, $p \leq 0.01$) (a, left), and splicing in *TET2*-mutant samples was compared to WT (MUT vs WT change ≥ 10%, $p \leq 0.01$) (a, right). Alternate splicing events were plotted as a proportion of total. Directional changes in inclusion reads were quantified for each category of alternate splicing event (b). Genes with alternately spliced exons that are hypermethylated in *TET2*-mutant samples and hydroxymethylated in *TET2*-wildtype samples are listed in (c).

Discussion

In aggregate, our data show that inactivating *TET2* mutations in MDS results in a global increase in DNA methylation, and that hypermethylation due to loss of *TET2* is enriched at intron-exon boundaries. Furthermore, we find that increased methylation at sites of alternate splicing correlates with shifts in splice variant ratios across a broad set of genes. These results are highly suggestive, given that DNA methylation at intron-exon boundaries affects mRNA splicing by altering interactions with RNA polymerase II binding partners (Flores et al., 2012; Gelfman et al., 2013; Shukla et al., 2011) and in light of the coincidence of somatic *TET2* and spliceosome mutations, and the effects of loss of *TET2/Tet2* on hematopoietic differentiation. We hypothesize that alterations in splice variant abundance due to aberrant methylation in *TET2*-mutant cells may affect hematopoietic differentiation—resulting in both an increase in self-renewal in the HSC compartment, and a monocyte-biased differentiation pattern—and promote the development of an MDS phenotype. These findings reveal a novel role for *TET2* in both normal and dysregulated hematopoiesis, and highlight the need for further research on the function of intragenic DNA methylation.

Chapter 4: Results – Part 2

TET2 Mutations Predict Response to Hypomethylating Agents in Myelodysplastic Syndrome Patients

Contributions

I developed protocols for *in vivo* and *in vitro* treatment with 5-azacitidine. In collaboration with others (see below), I performed all *in vivo* murine experiments, including drug treatment, competitive transplantation, peripheral blood analysis, and endpoint analysis.

All clinical study design was done by Rafael Bejar and David Steensma, and statistical analysis was performed by Rafael Bejar and Kristen Stevenson (DFCI Biostatistics Core). Clinical sample collection and data curation was done by Michal Bar-Natan, Richard M. Stone, Guillermo Garcia-Manero, Hagop Kantarjian, and David Steensma. Michelle Chen, and Marie McConkey provided technical support for *in vivo* murine experiments.

Note: The work presented in this chapter has been published (Bejar et al., 2014).

Summary

The molecular basis of MDS is still being unraveled, and targeted therapies are not yet available. Use of hypomethylating agents such as azacitidine or decitabine, however, has long been considered effective treatment (Garcia-Manero, 2008; Stresemann and Lyko, 2008). Mutations in TET2 are among the most common genetic lesions in MDS to date, and are presumed to lead to genomic hypermethylation. For this reason, we hypothesized that patients with TET2 mutations might be more sensitive to hypomethylating agents than patients with intact TET2.

Introduction

DNA hypomethylating agents are the only class of drugs approved for the treatment of patients with higher risk myelodysplastic syndromes (MDS). Azacitidine (AZA) was approved by the FDA for MDS in 2004 and later shown to confer an overall survival benefit compared to supportive care in a randomized phase-III study (Fenaux et al., 2009). Decitabine (DEC), the deoxynucleotide analog of AZA, was approved for the treatment of MDS in 2006 based on its ability to improve blood counts and decrease bone marrow blasts proportions (Steensma et al., 2009). However, only 40-50% of patients treated with either AZA or DEC experience hematologic improvement with these agents, and complete responses occur in as few as 10-15% of treated patients (Fenaux and Adès, 2009; Garcia-Manero, 2008). Effective methods for identifying patients who are most

likely to respond to treatment with a hypomethylating agent (HMA) would be of immediate clinical utility. Clinical features and patient characteristics may help stratify patients according to their response rates, but these models are not sufficiently conclusive to deny eligible patients a trial of AZA or DEC based on their predictions alone (Itzykson et al., 2011b; Zeidan et al., 2014). Better biomarkers of response to hypomethylating agents are needed.

Since the FDA approval of AZA and DEC, our understanding of the molecular genetic basis for MDS has expanded dramatically. Recurrent somatic mutations have been identified in more than 40 genes, and many of these mutated genes have been associated with important clinical measures including overall survival (Bejar et al., 2011; Haferlach et al., 2013; Papaemanuil et al., 2013). Since mutated genes underlie the pathogenic mechanisms driving the initiation and progression of MDS, they may represent molecular biomarkers of drug sensitivity or resistance. This is exemplified by the observation that MDS with deletions of the long arm of chromosome 5 (del[5q]) have a striking sensitivity to lenalidomide, whereas MDS patients without this lesion are less likely to have a hematologic response, and much less likely to have a cytogenetic or prolonged response (List et al., 2006). No such cytogenetic correlate has been found for the hypomethylating agents, but single gene mutations involving the pathways targeted by these drugs may be better candidates.

DEC and AZA (which is metabolized into DEC intracellularly), inhibit DNA methyltransferases and decrease the methylation of cytosine residues. Several of the most

frequently mutated genes in MDS encode proteins involved in the epigenetic regulation of gene expression such as TET2, DNMT3A, and ASXL1. DNMT3A is a *de novo* DNA methyltransferase and is a potential target of the hypomethylating agents. Somatic mutations of DNMT3A have been shown to decrease its activity, suggesting that pharmacologic targets other than DNMT3A are likely mediators of response to AZA and DEC (Stresemann and Lyko, 2008). Loss of function mutations in *TET2* impair the ability of this enzyme to oxidize methylcytosine residues and are associated with altered DNA methylation patterns and decreased 5-hydroxymethylcytosine (5hmC) levels in MDS patient samples (Ko et al., 2010; Yamazaki et al., 2014). A small study of AZA-treated MDS patients using Sanger sequencing to determine the mutation status of *TET2* found that mutations of this gene were associated with a slightly higher rate of response than in wild-type *TET2* patients (Itzykson et al., 2011a). However, the investigators did not examine these samples for additional mutations that might have modified this result and did not use techniques sensitive enough to identify mutations in small disease subclones. Subclonal mutations in genes associated with an adverse prognosis, including *ASXL1, RUNX1,* and *NRAS*, have been shown to have clinical relevance regardless of their abundance within the dysplastic clone (Murphy et al., 2013; Papaemanuil et al., 2013). These adverse mutations are often associated with disease progression and may mitigate the value of a sensitizing abnormality if they confer resistance to treatment.

We hypothesize that mutations of individual genes may serve as biomarkers of response for MDS patients treated with hypomethylating agents. We utilized massively parallel sequencing to examine 40 recurrently mutated genes in disease samples from 213

MDS patients treated with AZA or DEC. We examined the association of mutational patterns at different mutant allele fractions with response to treatment and overall survival. We utilized a competitive murine bone marrow transplant model to test the sensitivity of *Tet2*-null and *Tet2*-WT hematopoietic cells to treatment with azacitidine.

Results

Spectrum of Mutations

We examined tumor samples collected from 213 patients from three different sites prior to treatment with AZA, DEC, or DEC + another agent. There were no significant differences in pretreatment patient characteristics by treatment site (**Table 4.1**) or baseline characteristics as shown in **Table 4.2**. Frequently mutated regions of 40 genes previously shown to be somatically mutated in patients with MDS were subject to hybrid capture and massively parallel sequencing (**Supplemental Table 4.1**). These include the most frequently mutated splicing factors, kinase signaling genes, transcription factors, and epigenetic regulators such as *TET2*, *DNMT3A*, *ASXL1*, and *EZH2*. With this panel, we identified one or more mutations in 39 genes (**Figure 4.1**). In total, 94% of patients had a mutation in at least one recurrently mutated gene. The most frequently mutated genes were *ASXL1* (46%), *TET2* (27%), *RUNX1* (20%), *TP53* (18%), and *DNMT3A* (16%) followed by the splicing factor genes *SRSF2* (16%), *SF3B1* (15%), and *U2AF1* (14%).

Table 4.1: Patient Characteristics and Treatments Received by *TET2* Mutational Status

	N (%)	*TET2* wt	*TET2* Mut	p-value†
N	213	155	58	
Treatments Received				
Azacitidine Alone	42 (20)	30 (19)	12 (21)	0.60
Decitabine Alone	144 (68)	103 (66)	41 (71)	
Decitabine + Other	27 (13)	22 (14)	5 (9)	
Age, ≥ 70 yrs.	103 (48)	72 (46)	31 (53)	0.44
Sex				
Male	155 (73)	118 (76)	37 (64)	0.085
Female	58 (27)	37 (24)	21 (36)	
FAB				
RA	30 (14)	19 (12)	11 (19)	0.13
RARS	24 (11)	15 (10)	9 (16)	
RAEB	125 (59)	97 (63)	28 (48)	
RAEB-t	6 (3)	4 (3)	2 (3)	
CMML	21 (10)	13 (8)	8 (14)	
Other	7 (3)	7 (5)	0 (0)	
IPSS				
Low	11 (5)	5 (3)	6 (10)	0.019
Int-1	86 (40)	56 (36)	30 (52)	
Int-2	76 (36)	61 (39)	15 (26)	
High	37 (17)	31 (20)	6 (10)	
Unknown	3 (1)	2 (1)	1 (2)	
Cytogenetics				
Normal or –Y	107 (50)	68 (44)	39 (67)	0.022
Complex	51 (24)	45 (29)	6 (10)	
-7/7q- isolated or	14 (7)	12 (8)	2 (3)	
+8 isolated	11 (5)	8 (5)	3 (5)	
20q- isolated	7 (3)	6 (4)	1 (2)	
5q- isolated or +1	3 (1)	3 (2)	0 (0)	
Other	13 (6)	9 (6)	4 (7)	
Unknown	7 (3)	4 (3)	3 (5)	

†Test includes only known categories, chi-square test used for cytogenetics.

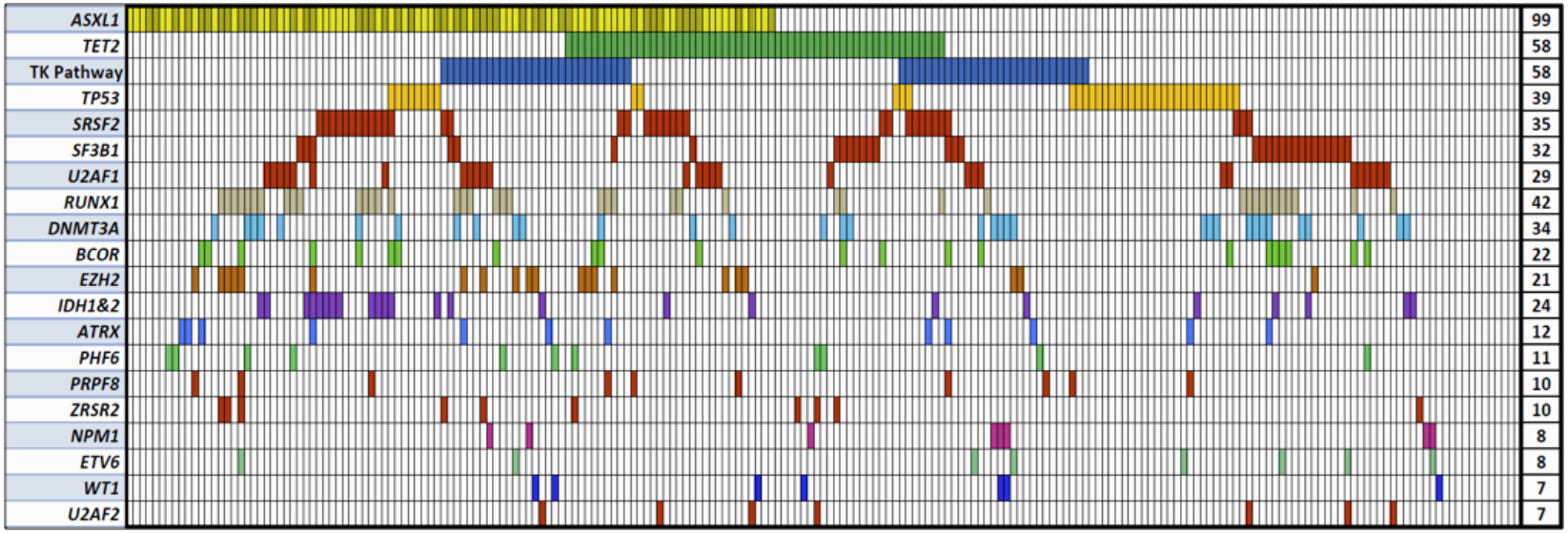


Figure 4.1: Spectrum of mutations in 213 patients in select MDS associated genes. Each column represents an individual patient sample and each colored cell represents a mutation of the gene or gene group listed to left of that row. The number of mutations for each row is indicated in the column to the right. Darker bars in the *ASXL1* row indicate patients with a p.G642fs mutation. TK Pathway = *NRAS*, *KRAS*, *CBL*, *CBLB*, *JAK2*, *PTPN11*, *BRAF*, *MPL*, *KIT*.

The frequency of mutations identified in these genes was largely similar to those identified in other MDS patient cohorts. Only *ASXL1* mutations were more frequent compared with prior studies, many of which examined a greater proportion of lower risk patients without transfusion dependence, utilized less sensitive Sanger sequencing of *ASXL1*, excluded unannotated missense mutations, or excluded insertions in a homopolymeric tract near amino acid 642 (Bejar et al., 2011; Haferlach et al., 2013; Papaemanuil et al., 2013). Other previously observed patterns of mutations were identified in this cohort including the paucity of *ASXL1* mutations in *SF3B1* mutant samples, the near mutual exclusivity of splicing factor mutations, and the lower rate of other gene mutations in patients with *TP53* mutations (**Supplemental Figure 4.1**) (Bejar et al., 2012; Yoshida et al., 2011). As expected, mutations of *TET2*, *ASXL1*, *NRAS*, *EZH2*, and *SRSF2* were overrepresented in CMML cases, *SF3B1* mutations were predominantly in RARS cases, and mutations of *TP53*, *IDH1*, and *IDH2* were relatively underrepresented in RA/RARS patients (**Supplemental Figure 4.1**). The variant allele fractions (VAFs) of mutations were not uniform and varied greatly for individual genes (**Figure 4.2**). For example, splicing factor abnormalities had higher median variant allele fractions, while the VAFs for mutations in tyrosine kinase signaling genes were lower, indicative of their frequent presence in disease subclones. Mutations of all genes included samples with variants present only at low abundance.

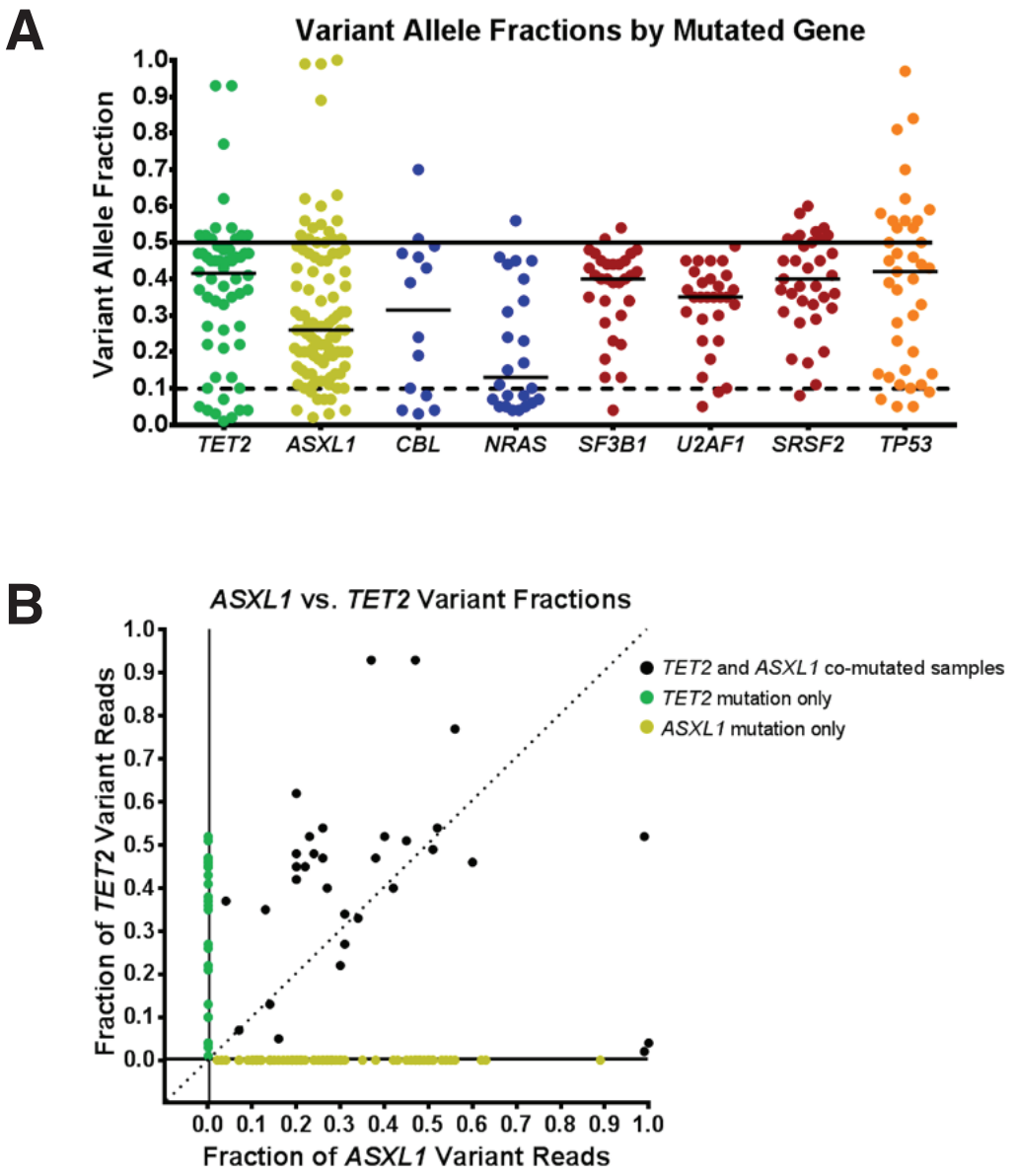


Figure 4.2: Variant Allele Frequencies in Selected Genes.
A) Quantitative measure of variant containing reads estimates the abundance of these mutations (uncorrected for allele copy number). Mutations of *TET2* (green), *TP53* (orange), and splicing factor genes (red) are often present in the dominant clone while mutations of tyrosine kinase signaling genes (blue) are often present in smaller clones. Mutations of *ASXL1* (yellow) are more widely distributed. B) Analysis of samples with both *TET2* and *ASXL1* mutations indicate that *ASXL1* mutations are most often co-dominant with or smaller than *TET2* mutant clone

Clinical Findings, Variant Allele Fraction, and Response Rates

The overall response rate of patients in the study was 47% with 31% achieving a CR according to IWG criteria revised in 2006 (**Table 4.2**). There was no significant difference in response by treatment regimen (p=0.96) or source of sample (p=0.36). IPSS risk groups and cytogenetic abnormalities were not associated with response rate. The only clinical feature significantly associated with response rate was FAB classification (p=0.008) driven largely by the high response rate of CMML patients (17/21, 81%). Thirty-five percent of RA/RARS patients achieved a response compared with 47% of RAEB patients. No differences were detected in the time to response for each mutation.

Table 4.2: Response vs. Pt. Characteristics and Treatment

	Total	Non-Responders N (%)	Responders N (%)	p-value†
N, Total	213	113 (53)	100 (47)	
Treatment				
Azacitidine Alone	42 (20)	22 (52)	20 (48)	0.96
Decitabine Alone	144 (68)	76 (53)	68 (47)	
Decitabine + Other	27 (13)	15 (56)	12 (44)	
Sex				
Male	155 (73)	82 (53)	73 (47)	0.99
Female	58 (27)	31 (53)	27 (47)	
Age				
<70 yrs.	110 (52)	64 (58)	46 (42)	0.13
≥70 yrs.	103 (48)	49 (48)	54 (52)	
FAB				
RA	30 (14)	20 (67)	10 (33)	0.008†
RARS	24 (11)	15 (63)	9 (38)	
RAEB	125 (59)	65 (52)	60 (48)	
RAEB-t	6 (3)	4 (67)	2 (33)	
CMML	21 (10)	4 (19)	17 (81)	
Other	7 (3)	5 (71)	2 (29)	
IPSS Risk Group				
Low/Int-1	97 (46)	53 (55)	44 (45)	0.78††
Int-2/High	113 (53)	59 (52)	54 (48)	
Unknown	3 (1)	1 (33)	2 (67)	
Cytogenetics				
Normal	104 (49)	49 (47)	55 (53)	0.31
Complex	51 (24)	28 (55)	23 (45)	
Other	51 (24)	31 (61)	20 (39)	
Unknown	7 (3)	5 (71)	2 (29)	

†Test includes only known categories.

††No difference was observed between the 4 individual IPSS categories (p=0.24).

In a prior study by Itzykson *et al.*, mutations of *TET2* detected by Sanger sequencing were found to predict a nearly two-fold greater response rate to AZA (Itzykson et al., 2011a). In our cohort, *TET2* mutant patients showed only a trend toward increased response rates compared to wild-type (WT) (55% vs. 44%; OR 1.58 [0.86-2.89], p = 0.14) and no other mutated was gene was associated with a significantly improved overall response rate in univariate analyses (**Table 4.3, Supplemental Table 4.2, Supplemental Figure 4.2**). However, the VAF for mutations of *TET2* and several other genes spanned a wide range of values including many likely to be below the detection limit for Sanger sequencing (**Figure 4.2**). We hypothesized that mutations capable of sensitizing cells to hypomethylating agents are more likely to be associated with a clinical response to treatment when they are present in a major disease clone. For example, even complete elimination of a clone representing less than 20% of bone marrow cells might not have any impact on the assessment of clinical response. Therefore, we repeated our analysis with mutations present at a VAF of <10% treated as if they were unmutated. For heterozygous mutations, this cutoff represents variants present in less than 20% of the sample cellularity and is at the estimated limit of sensitivity for Sanger sequencing. In this revised analysis considering VAF, mutations of *TET2* were associated with a significantly higher response rate compared to WT (60% vs. 43%; OR 1.99 [1.0.5-3.80], p = 0.036; adjusted OR 1.98 [1.02-3.85], p=0.044), and comparable to that shown by Itzykson *et al* (**Table 4.3, Supplemental Table 4.2, Supplemental Figure 4.2**).

Table 4.3: Association of gene mutations with response rate in logistic regression analysis

Mutated Gene†	Unadjusted OR (95% CI)	p-value	Adjusted†† OR (95% CI)	p-value
Mutations with VAF ≥ 10%				
TET2 mut vs. *TET2* wt	1.99 (1.05, 3.80)	0.036	1.98 (1.02, 3.85)	0.044
ASXL1 mut vs. *ASXL1* wt	0.69 (0.40, 1.20)	0.19	0.68 (0.38, 1.19)	0.17
TET2 mut + *ASXL1* wt vs. other	3.65 (1.38, 9.67)	0.009	3.64 (1.35, 9.79)	0.011
TET2 mut + *ASXL1 wt* vs. both wt	3.40 (1.24, 9.35)	0.011	3.36 (1.20, 9.38)	0.013
TET2 wt + *ASXL1* mut vs. both wt	0.77 (0.41, 1.46)	0.035	0.80 (0.39, 1.46)	0.039
TET2 mut + *ASXL1* mut vs. both wt	1.11 (0.48, 2.61)	0.62	1.07 (0.44, 2.61)	0.59
CBL mut vs. *CBL* wt	0.27 (0.06, 1.29)	0.10	0.28 (0.06, 1.40)	0.12
Including All Mutations				
TET2 mut vs. *TET2* wt	1.58 (0.86, 2.89)	0.14	1.60 (0.85, 3.02)	0.15
ASXL1 mut vs. *ASXL1 wt*	0.77 (0.45, 1.32)	0.34	0.74 (0.42, 1.30)	0.29
TET2 mut + *ASXL1* wt vs. other	2.37 (1.00, 5.58)	0.049	2.40 (0.99, 5.79)	0.051
TET2 mut + *ASXL1 wt* vs. both wt	2.27 (0.91, 5.63)	0.055	2.27 (0.89, 5.79)	0.056
TET2 wt + *ASXL1* mut vs. both wt	0.86 (0.45, 1.64)	0.16	0.84 (0.43, 1.62)	0.15
TET2 mut + *ASXL1* mut vs. both wt	1.06 (0.47, 2.38)	0.67	1.04 (0.45, 2.44)	0.68
CBL mut vs. *CBL* wt	0.17 (0.04, 0.79)	0.023	0.18 (0.04, 0.82)	0.027

†Reference group is listed second.
††Adjusted for sex, age (<70, ≥70 yrs.), IPSS (Low/Int1 vs. Int2/High), and treatment (azacitadine alone vs. decitabine alone vs. decitabine ± other), none of the Hosmer and Lemeshow tests indicated a lack of fit for each model.
wt=wild-type, mut=mutant

When all VAFs were considered, only mutations of *CBL*, which were often of low VAF, were associated with a lower rate of response compared to WT (14% vs. 49%) in this analysis (OR 0.17 [0.04-0.79], p=0.023; adjusted OR 0.18 [0.04-0.82], p=0.027), but was not significant when low VAF mutations were considered WT (20% vs. 48%) (OR 0.27 [0.06-1.30], p=0.10).

By sequencing multiple genes, we had the opportunity to determine if mutations in additional genes could modulate the response rates of *TET2* mutant patients. We focused on the subset of patients defined by their *TET2* and *ASXL1* mutation status (at any VAF) as these contained enough individuals for a meaningful statistical analysis. Patients with mutated *TET2* an unmutated *ASXL1* demonstrated an increased overall response rate compared to all others (65% vs. 44%; OR 2.37, [1.00-5.58], p=0.049) (Table 3). This effect was more pronounced when mutations were required to have a VAF ≥ 10% (74% vs. 44%; OR 3.65, [1.38-9.67], p = 0.009), representing over 10% of patients in this cohort (**Table 4.3**).

In vivo *Model of Azacitidine Response in Tet2$^{-/-}$ Cells*

The observed association between *TET2* mutations and response to treatment could be mediated directly by *TET2* loss-of-function or by indirect or cell extrinsic effects. To test whether TET2 loss-of-function can sensitize cells to hypomethylating agents, we performed a competitive murine bone marrow transplant experiment using hematopoietic cells from *Tet2*-null and wildtype littermate donors. As expected, equal numbers of CD45.2^{+} cells transplanted into CD45.1^{+} recipients resulted in greater

engraftment of *Tet2*-null cells by 2.5 weeks post-transplant, prior to treatment with azacitidine. There was no difference in peripheral blood counts between groups at this time point. Treatment with AZA (2.5 mg/kg M-F x 2 weeks) or vehicle was begun on day 20 post-transplant and repeated starting on days 48, 76, and 104. Regardless of genotype, azacitidine-treated animals exhibited significant decreases in WBC levels and hematocrit and an initial drop in peripheral blood chimerism. For several subsequent cycles, AZA-treated *Tet2*-null cells maintained a significantly decreased representation in peripheral blood while *Tet2*-wildtype cells did not (**Figure 4.3**).

Figure 4.3: Peripheral blood chimerism over time is shown after competitive bone marrow transplantation with cells from 45.2 *Tet2*-null mice in panel A and 45.2 *Tet2*-wildtype mice in panel B. Gray bars indicate periods of treatment with AZA or vehicle. *Tet2*-null cells show increased chimerism compared with *Tet2*-wildtype cells. Treatment with azacitidine significantly decreases chimerism in the *Tet2*-null recipient mice only. * $p < 0.05$, ** $p < 0.01$

Associations with Overall Survival

Traditional prognostic models like the IPSS and Revised IPSS (IPSS-R) are based on patient cohorts examined only until they receive disease modifying therapies such as hypomethylating agents or undergo stem cell transplantation. Response to specific treatments could significantly alter the prognostic impact of adverse disease features or genetic alterations. We explored the relationship between mutation status and overall survival in the subset of patients with available survival data. Of these 146 patients (69%) in our cohort, 119 died during follow up. The median follow-up for patients remaining alive was 3.8 years (95% CI, 3.1-5.8). Despite its association with response, *TET2* mutation status was not associated with overall survival (p=0.56), consistent with the finding in Itzykson *et al* (**Figure 4.4A**). Mutations of *TP53* were associated with lower overall survival (21% of patients; HR 2.01 [1.29-3.14], p=0.002; adjusted HR 1.91 [1.20-3.05], p=0.007; **Figure 4.4B**) as were the much rarer mutations of *PTPN11* (4% of patients, HR=3.26 [1.41-7.58], p=0.006; adjusted HR 2.47 [0.98-6.26], p=0.056; **Figure 4.4C**)].

Overall survival in patients with complex karyotypes was strongly associated with *TP53* mutation status (**Figure 4.4D**). Patients with both a complex karyotype and a *TP53* mutation had a median survival of only 0.9 years. In contrast, patients with complex cytogenetics and no *TP53* mutation had an overall survival of 1.3 years which was not different from patients with karyotypes other than complex (median 1.8 years, *p*=0.28). This indicates that the adverse prognostic value ascribed to the complex karyotype is

largely driven by its frequent association with *TP53* mutations which could be used to better refine disease risk in this patient population.

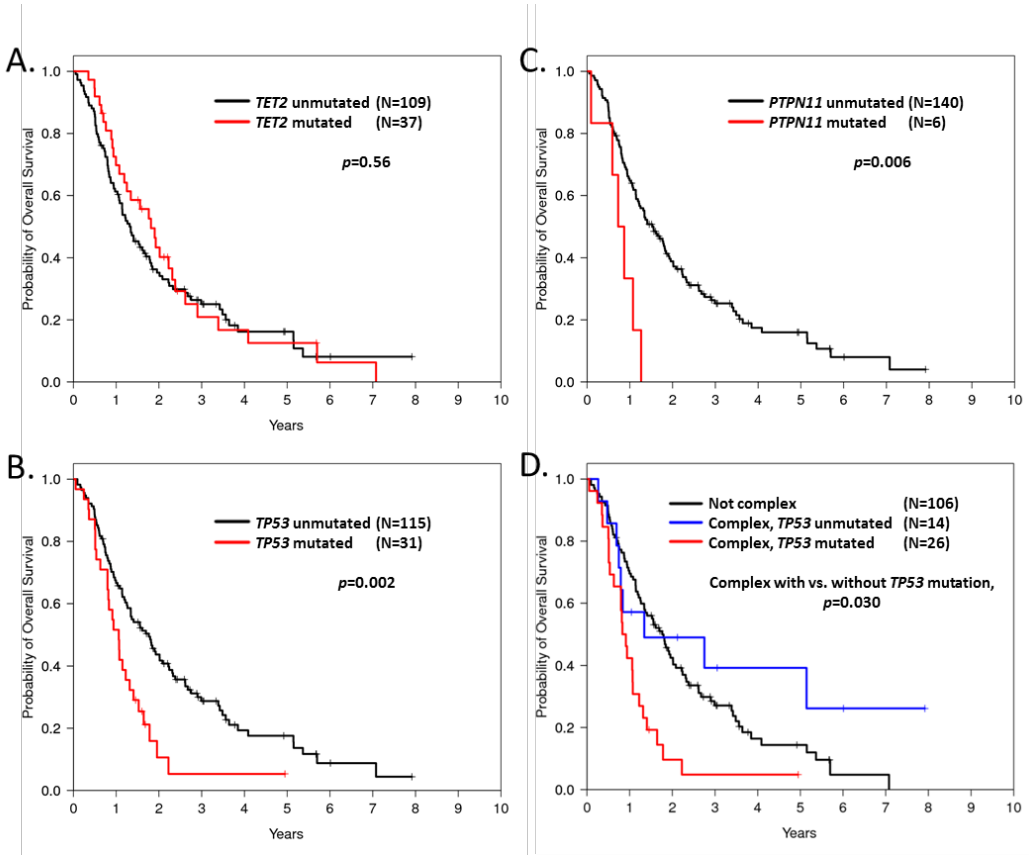


Figure 4.4: Kaplan-Meier curves for overall survival in the 146 out of 213 study patients with survival data. A) Survival of patients with and without *TET2* mutations B) Survival of patients with and without *TP53* mutations. C) Survival of patients with and without *PTPN11* mutations. D) Survival of complex karyotype patients with and without *TP53* mutations vs. patients without complex karyotypes.

Discussion

In our study, the presence of *TET2* mutation at greater than 10% allele burden predicted an increased response to hypomethylating agents, particularly in the subset that lacked similarly abundant mutations of *ASXL1*. To achieve this result, we examined tumor samples from 213 MDS patients collected prior to treatment with hypomethylating agents for mutations in 40 genes known to be recurrently mutated in MDS. The patients in our cohort were representative of those studied in clinical trials of AZA and DEC in terms of predicted disease risk and severity of cytopenias. Overall response rates were just under 50% and did not differ by the type of drug patients received. Using sensitive quantitative sequencing techniques, we were able to identify mutations in over 90% of patients in patterns similar to those seen in prior multigene studies of MDS.

Our findings are consistent with those of Itzykson *et al.*, who previously reported that 11 of 13 (85%) MDS patients with *TET2* mutations detected by Sanger sequencing responded to treatment with AZA compared with a 52% response rate for their overall cohort of 86 patients. In that study, mutations in other genes were not examined, and small subclonal *TET2* mutations likely went undetected. Our broader and more sensitive multigene analysis similarly identified *TET2* mutations as predictive of response to hypomethylating agents in a larger cohort of patients. Surprisingly, consideration of mutations in other genes did not reveal additional predictors of favorable response and inclusion of low VAF mutations weakened the association between *TET2* mutation status and response rate. However, our approach identified the 10% of patients with mutated

TET2 and wildtype *ASXL1* as the group most likely to respond to treatment. Potential explanations for this finding include partial resistance to hypomethylating agents caused by *ASXL1* mutations. In this model, the *ASXL1* mutated subclone would be expected to grow in size during disease progression or relapse and might confer primary resistance. *ASXL1* mutated patients with wildtype *TET2* did have a lower likelihood of response, but this was not statistically significant (OR 0.63 [0.35-1.15], p=0.13). Alternatively, we observed that *ASXL1* mutations were often subclonal or at a lower VAF than *TET2* mutations in co-mutated patients (**Figure 4.2C**). The acquisition of such secondary mutations (of which *ASXL1* was the most frequent) could indicate more clonally progressive disease that might be inherently more resistant to treatment.

The mechanism by which *TET2* mutations might influence response to hypomethylating agents is not clear. Altered methylation has been observed in patients with *TET2* mutations and in animal models of *Tet2* loss. However, measurement of pretreatment DNA methylation by itself has not been found to be predictive of response to hypomethylating agents (Shen et al., 2010). In our murine bone marrow transplant experiment, exposure to AZA preferentially decreased the clonal advantage associated with loss of Tet2 function. This effect may be associated with a greater AZA sensitivity in more actively cycling cells since AZA results in cell division-dependent passive demethylation of DNA. Mice with hematopoietic *Tet2* loss are known to have increased myeloid progenitor proliferation (Abdel-Wahab et al., 2012; Kunimoto et al., 2012; Li et al., 2011; Moran-Crusio et al., 2011).

An important finding of our study was that no pattern of mutation was strongly associated with a lack of response to treatment. Responses to hypomethylating agents were observed even in patients with mutations that confer a very poor prognosis. Therefore, our data indicate that mutation information alone should not be used as a basis for denying therapy with an HMA if treatment is indicated. Studies examining samples collected at multiple time points are needed to identify mutations predictive of acquired resistance or relapsed disease (Jadersten et al., 2009; Walter et al., 2013; 2012).

The association between molecular or clinical biomarkers and HMA response may be confounded by the by variations in enzymes responsible for the activation and metabolism of AZA and DEC (Mahfouz et al., 2013; Saunthararajah, 2013; Stresemann and Lyko, 2008; Valencia et al., 2014). Patients who demonstrate significant hypomethylation of blood cell DNA after treatment with AZA or DEC (indicating sufficient exposure to target DNA methyltransferases) may be more likely to have a clinically significant response (Qin et al., 2011; Shen et al., 2010). It is possible that the predictive value of cell intrinsic somatic mutations may be enhanced if controlled for cell extrinsic variables such as effective dose of HMAs received and activated in cells.

Over two-thirds of our cohort had survival data collected. Mutation profiles capable of predicting response to hypomethylating agents were not associated with differences in overall survival (**Figure 4.3**). However, mutations in two genes that were not predictive of response, *TP53* and the rarer *PTPN11*, were each associated with decreased overall survival. The majority of *TP53* mutant patients had a complex

karyotype, a known adverse risk factor associated with shorter overall. More than half of our complex karyotype patients harbored a *TP53* mutation (32/51) and these patients had a very short overall survival (median 0.9 years). However, complex karyotype patients without a detectable *TP53* mutation had an overall survival that was no different from the group of patients with non-complex karyotypes. This indicates that the negative prognostic significance attributed the complex karyotype can be better explained by the *TP53* mutation status of these patients and validates the results of recent studies in MDS and AML (Bally et al., 2014; Bejar et al., 2011). In addition to *TP53*, mutations in any of four additional genes, *RUNX1*, *ASXL1*, *EZH2*, or *ETV6* were found to predict shorter overall survival than expected by examining clinical features alone. However, mutations of these four genes were not found to be prognostically adverse in this cohort of treated patients (**Supplemental Figure 4.3**). In contrast, samples for our previous study were collected prior to the approval of AZA and DEC and therefore came largely from untreated patients. Our results suggest that treatment with hypomethylating agents may partially abrogate the adverse prognostic impact of these lesions. If validated, our finding would form a justification for treating patients whose adverse prognosis is driven by mutations in these genes.

The clinical implications of our findings are that response to hypomethylating therapy can be predicted in a subset of patients using molecular genetic features. A more robust predictor might be created by incorporating clinical findings or other biomarkers (Alhan et al., 2014; Itzykson et al., 2011b; Valencia et al., 2014). Indeed, the Groupe Français des Myélodysplasies has presented a clinically and cytogenetically based

prognostic model for azacitidine treated patients, although its predictive power is unclear (Breccia et al., 2012; Itzykson et al., 2011b; Zeidan et al., 2014). As with these clinical measures, no mutations identified in our study were reliably strong predictors of primary resistance to treatment in a large number of patients. Therefore, there is no genetic rationale for denying MDS patients the opportunity to be treated with AZA or DEC based on our findings, particularly since there are few alternative therapies approved for this patient population.

In conclusion, means of reliably predicting response to hypomethylating agents would be of clinical benefit in the care of patients with MDS. Our study demonstrates that mutation profiles can help in this effort to some extent. Studies examining the mechanism by which these biomarkers might mediate sensitivity or resistance to treatment would be of clinical value and could lead to the discovery of additional therapeutic targets in MDS.

Chapter 5

Discussion

Summary

I have shown that loss of *TET2* in primary MDS samples leads to global hypermethylation, and that this increase in methylation is localized to intron-exon boundaries. I have developed a TF1 cell line system to assess changes in hydroxymethylation and transcription in a *TET2*-null setting. Using this system, I have shown that loss of *TET2* results in a global loss of 5-hmC, and that hypermethylated regions in *TET2*-mutant samples are enriched for 5-hmC-containing regions in *TET2*-wildtype cells. Loss of *TET2* has little impact on global transcription, but does affect alternate splicing events: exon-skipping, as assessed by shifts in cassette, coordinate cassette, and mutually exclusive exon categories of alternate splicing events, is increased in *TET2*-null cells. Regions hypermethylated in *TET2* mutant samples are enriched for alternate splicing events overall, suggesting that exon skipping is directly due to increased methylation seen in the *TET2*-deficient setting. Finally, I have shown that murine *Tet2*-null cells display increased sensitivity to treatment with 5-azacitidine, functionally validating our findings that MDS patients with *TET2* mutations show improved response to treatment with hypomethylating agents.

Implications and Future Directions

TET proteins and Intragenic Methylation

DNA methylation has been best studied in the context of gene promoters, and, more recently, enhancers. Because of this, methylcytosine has long been thought of as a

repressive mark. The work I have presented in this thesis suggests that the reality of methylcytosine-mediated control is more complicated.

Loss of Tet1 in murine ES cells has a dual effect on gene expression, and Tet1 binds to both actively transcribed and repressed promoters (Wu et al., 2011). Several groups have shown that Tet1-mediated demethylation is important for the derepression of some genes—notably pluripotency factors such as Nanog—but may actually directly or indirectly repress expression of other genes (Pastor et al., 2011; Williams et al., 2011; Wu et al., 2011). These results indicate that TET1 function is closely linked to promoter-focused regulation.

Gains in methylation due to loss of TET2 in adult human hematopoietic cells, unexpectedly, is not localized to promoters or enhancers. My data show, rather, that changes in methylation in a TET2-deficient setting are found primarily in intragenic regions, and predominate at intron-exon boundaries. Given the effects of *TET2* loss on hematopoietic cells, these results suggest that intragenic methylation plays an important functional role in self-renewal and differentiation decisions. These results also suggest that TET2 function in hematopoietic cells may not overlap with that of TET1 in ES cells.

It will be important to further study the function of intragenic methylation, particularly at intron-exon boundaries. Several groups have described a role for intragenic methylation in modulating mRNA splicing, primarily by permitting or interfering with binding of RNA polymerase II (Pol II) interacting proteins (Flores et al., 2012; Gelfman et al., 2013; Maunakea et al., 2013; Shukla et al., 2011). In this context, DNA methylation has universally been shown to affect exon inclusion or exclusion wholesale, rather than altering choice of splice sites. Interestingly, reports are conflicted as to

whether increased methylation of a given exon promotes its inclusion or exclusion, which likely reflects some degree of tissue- or cell-type-specificity in the expression of Pol II interacting proteins.

Targeted studies of methylated DNA in the context of transcription-translation are technically challenging, and previous work has focused on association studies (Flores et al., 2012; Gelfman et al., 2013), experiments on known Pol II-interacting proteins (Shukla et al., 2011), or genome-wide manipulations of 5-mC content through DNMT1 depletion or treatment with HMAs (Maunakea et al., 2013). These last treatments are blunt tools, however, and may have indirect effects on transcription or translation more broadly. One possible approach to studying the effects of methylation of a single exon on its inclusion in mature mRNA and protein would be to use a minigene system for coupling transcription, splicing, and translation (Cooper, 2005; Rothrock et al., 2003; Singh and Cooper, 2006).

A genomic region of interest, including three exons as well as their intervening introns, would need to be cloned behind a strong promoter in either a plasmid or as linear DNA. Appropriate restriction enzyme sites would need to be identified surrounding the middle exon, and this middle fragment could be PCR amplified using a custom dNTP mix containing 5-mC rather than C. This would result in a fully-methylated middle exon, which could be ligated back into its original genomic context. The resulting plasmid or linear DNA could be incubated directly with cellular extract from the cell type of interest (ideally from both *TET2*-deficient and *TET2*-wildtype cells), and *in vitro* transcription of the minigene measured by RT-PCR using plasmid-specific primers.

This approach would obviously be heavily dependent on choice of promoter as well as choice of genomic region—intron size would likely be a major technical consideration—and is not amenable to high-throughput studies. A minigene is inherently artificial, but this approach would demonstrate a direct link between methylation and splicing. This approach also has the added benefits of biological relevance and flexibility as to the available pool of possible Pol II interacting proteins, which can be varied according to choice of cellular extract.

Finally, it would be interesting to interrogate the pool of Pol II binding proteins present in hematopoietic cells, and how those interactions are affected by loss of *TET2* and corresponding gains in methylation in a cell-specific context. An initial approach could involve immunoprecipitation of Pol II in *TET2*-null and *TET2*-wildtype hematopoietic cells, followed by immunoblotting for known Pol II interacting proteins. This approach would be unable to identify novel Pol II interacting proteins, however.

An unbiased approach to identify differential protein binding in the context of *TET2* loss would be to use SILAC followed by mass spectrometry to identify DNA binding proteins in a region of interest (Mittler et al., 2009; van der Ploeg, 2009). Biotinylated DNA oligos, with sequence derived from exons known to be hypermethylated and skipped in the context of loss of *TET2*, would be used as bait. *TET2*-null and *TET2*-wildtype TF1 cells (or another cell type of interest) would be grown in both heavy and light media, and nuclear extracts from these four conditions would be prepared and incubated with bait, which would then be pulled down and analyzed by liquid chromatography MS/MS. A control bait, with altered sequence, would need to be used to eliminate non-specific binding. Ultimately, proteins that bind to the sequence of

interest regardless of *TET2* mutational status could be identified, but a second population of proteins that preferentially bind to the sequence of interest in the unmethylated (bait) state would be most interesting. In *TET2*-wildtype cells, bait sequence would compete for binding of these proteins with unmethylated genomic DNA, whereas in *TET2*-mutant cells these proteins would be unable to bind genomic DNA due to methylation status, and should be enriched for after pulldown. The limiting factor to this approach, aside from cost, would be the bait sequence—sequences of interest would need to be carefully designed, and this approach probably does not lend itself to a high-throughput screen.

The interactions of any of proteins identified using this approach with the sequence of interest, as well as potentially with Pol II, would need to be validated using standard immunoprecipitation methods. Finally, to assess the role of protein-DNA binding on exon skipping, it would be interesting to disrupt the binding of candidate proteins (preferably by inducing targeted mutations in DNA-binding domains, though depletion of candidate proteins by shRNA or genetic deletion are alternate approaches) to DNA in a *TET2*-wildtype context.

Functional Significance of Hydroxymethylcytosine

DNA methylation increases in a site-specific manner with loss of TET2, and we also see a concomitant loss of 5-hmC. As discussed previously, 5-hmC may be a transient intermediary in the demethylation process, but it also may be a functional mark in its own right. Although methylcytosine content can vary depending on tissue type (Ehrlich et al., 1982), hydroxymethylcytosine can be found at much more variable frequencies in different tissues (Globisch et al., 2010; Kinney et al., 2011; Nestor et al., 2012; Song et

al., 2011). This may point to higher turnover of mC, indicating a greater need for active demethylation in certain tissue types, but it could also indicate that 5-hmC is functionally important in some tissues.

Work by several groups has shown that some MBD proteins can also bind 5-hmC, and a recent screen identified several proteins that preferentially bind to hydroxymethylated DNA (Iurlaro et al., 2013; Pastor et al., 2013). Furthermore, some MBD proteins cannot bind to 5-hmC, suggesting that 5-hmC might be used to fine-tune protein-DNA interactions (Pastor et al., 2013). Further work will be needed to dissect the functional significance of the interactions of these proteins with modified cytosine. Tet1 and 5-hmC are both enriched for at poised, or bivalent (marked by both activating H3K4Me3 and repressive H3K27Me3), promoters in ES cells, suggesting that 5-hmC may play a role in regulating stem cell maintenance and/or differentiation decisions (Pastor et al., 2011). The significance of intragenic 5-hmC has not been explored, although several groups have shown that 5-hmC is enriched for in gene bodies (Song et al., 2011; Tsagaratou et al., 2014).

Hydroxymethylcytosine is a rare base, comprising approximately 0.1% of total cytosine in most tissues. The most tractable system for studying 5-hmC is probably the brain, where 5-hmC is most abundant and makes up 0.3-0.6% of total cytosine (Globisch et al., 2010; Kriaucionis and Heintz, 2009). There may be tissue-specific differences in 5-hmC-binding proteins, and the activity of TET proteins may vary depending on tissue type, so any interactions found in this system would need to be verified in others. Several approaches mentioned above could be used to study protein-hmC interactions, including an examination of interactions between known 5-hmC-binding proteins with Pol II in the

context of wildtype versus mutant *TET*, an unbiased proteomics-based approach to identifying novel 5-hmC interacting proteins, and minigene approaches to assess effects of 5-hmC on cotranscriptional splicing.

Other TET functions

The focus of my thesis has been on the role of TET2 as a methylcytosine dioxygenase, and the effects of its loss on DNA methylation and hydroxymethylation. Recent work has shown that TET family proteins can interact with OGT and the SET1/COMPASS histone methyltransferase complex (Deplus et al., 2013; Vella et al., 2013). The possibility exists that the loss of this interaction, in addition to loss of methylcytosine dioxygenase activity in hematopoietic cells, may explain some of the phenotypes seen with loss of TET2. It will be important to investigate how these interactions affect hematopoietic specification and differentiation.

TET colocalization with OGT has been described in murine ES cells and HEK293 cells as occurring primarily at promoters, however (Deplus et al., 2013). This interaction may therefore be functionally distinct from interactions resulting in intragenic hydroxymethylation. It would be interesting to examine intragenic methylation/hydroxymethylation, as well as TET2 localization, in the context of OGT depletion in hematopoietic cells. For initial *in vitro* studies, primary mouse or human cells would be most relevant, but a hematopoietic lineage cell line with wildtype TET2, such as the TF1 cell line, might also be a tractable experimental system. *In vivo* studies could be performed in the conditional Ogt^{F} mouse crossed to a line expressing hematopoietic lineage-specific Cre (e.g. Vav-Cre or Mx-Cre) (Shafi et al., 2000).

Alternate Splicing

Our data shows an increase in exon skipping events with loss of TET2, and that alternate splicing events are enriched for in hypermethylated regions. As discussed above, the importance of DNA methylation in relation to mRNA splicing will be an interesting area of future research. The role of alternate splicing more broadly, in MDS and other hematologic malignancies as well as in normal hematopoiesis, will also be important to study. This is especially true in light of the high frequency of mutations in splicing factors in myeloid malignancies.

Expression of alternate mRNA isoforms are known to affect both normal development and disease. Hematopoietic differentiation, both developmental and adult, can be profoundly affected by altering expression of isoforms of *Runx1* (*AML1*), for example. Aberrant expression of Runx1a, which acts in a dominant negative manner to inhibit Runx1b function, leads to a loss of HSC self-renewal and altered differentiation, and promotes leukemogenesis (Liu et al., 2009). Several inherited disorders are known to be caused by *cis*-acting mutations that affect mRNA splicing, usually by altering splice sites, and alterations in the ratios of isoform expression can also lead to disease. In Parkinson's disease, for example, mutations in the *MAPT* (Tau) gene promote the inclusion of exon 10, leading to an increase in the abundance of an isoform of Tau with higher microtubule binding affinity, which in turn promotes Tau aggregation. Similarly, mutations affecting expression of different tumor suppressor or oncogene isoforms have been shown to play a role in the development of various cancers. (Douglas and Wood, 2011)

Approaches to studying the role of splicing in hematopoiesis and hematologic malignancy are clearly needed. It will be important to consider the expression of different mRNA isoforms, rather than exclusively gene-level changes, when performing transcriptional analyses. A shift from microarray-based approaches, which are limited in their ability to detect known splice variants and cannot typically detect novel splice variants, towards unbiased sequencing of the transcriptome is an important step. A current lack of consensus on the most appropriate methods for analysis of alternate splicing events in RNA-sequencing data is another obstacle that will need to be resolved. As alterations in splicing are discovered in MDS and other hematologic malignancies, a straightforward way to explore their functional relevance will be to overexpress specific isoforms, but dosage effects will need to be taken into consideration. This approach lends itself best to analyzing the role of changes in expression of mRNA isoforms within a single gene or set of target genes—the phenotypic effects of widespread changes in transcription may prove more challenging to dissect.

MDS Therapies

We have shown that MDS patients with *TET2* mutations respond significantly better to treatment with HMAs than do patients without the same, or patients with more complicated combinations of genetic lesions. Response in these cases was measured as improvements in clinical parameters, but not on survival, however. Our data indicated that while treatment with azacitidine abrogated the repopulation advantage in *Tet2*-null cells, this effect was at the level of a progenitor or ST-HSC, and LT-HSCs remained

unaffected. Treatment with HMAs is therefore unlikely to eradicate disease, though it may help to control proliferation of more differentiated cells.

Since *TET2* mutations frequently co-occur with other somatic mutations, it would be interesting to test combination therapies of HMAs and other targeted therapies. Given the frequent coincidence of *TET2* mutations and mutations in *SRSF2* or *ZRSR2*, it would be especially relevant to combine HMAs and spliceosome inhibitors in the hopes of finding a therapeutic window for effective treatment. Human CD34+ cells from MDS or AML patients are more sensitive to SF3B1 inhibitors than are wildtype CD34+ cells (Balaian et al., 2013a; 2013b). Other drugs targeting wild-type splicing factor function exist (Bonnal et al., 2012), but the development of compounds targeting the neomorphic activies of mutant splicing factors would be incredibly useful. A chemical screen using a methylation-independent version of the minigene approach described above coupled with a luciferase or GFP reporter system could be used to sensitively detect changes in splicing activity in a relatively high-throughput fashion.

Approaches to testing the efficacy of combination therapy could be done *in vitro* using patient-derived CD34+ cells, or *TET2*-deficient and splicing factor-neomorphic cell lines, but our experience has shown that the effects of azacitidine treatment are best modeled *in vivo*. Existing conditional mouse models for *Tet2* and *Srsf2* loss, in a background expressing hematopoietic lineage-specific Cre, would be an ideal model for preclinical studies. Competitive transplants allowing for the establishment of an engraftment advantage or disease phenotype represent an attractive system for assessing the effects of therapy on all hematopoietic compartments.

Concluding Remarks

This work has demonstrated a novel role for TET2 in specifically modulating intron-exon DNA methylation, and raises questions about the functional role of methylation in this region in hematopoiesis, myelopoiesis, and the development of myeloid disease. The study of DNA methylation has expanded considerably over the past decade, and it will be extremely interesting to see how the field resolves the challenges of integrating methylation and splicing data. Finally, while the treatment options available for MDS remain limited, our increased understanding of the molecular basis of the disease should allow for increased tailoring of treatment to individual patients, as well as combination therapy to more effectively target the MDS clone.

References

Abdel-Wahab, O., Adli, M., LaFave, L.M., Gao, J., Hricik, T., Shih, A.H., Pandey, S., Patel, J.P., Chung, Y.R., Koche, R., et al. (2012). ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression. Cancer Cell *22*, 180–193.

Abdel-Wahab, O., Manshouri, T., Patel, J., Harris, K., Yao, J., Hedvat, C., Heguy, A., Bueso-Ramos, C., Kantarjian, H., Levine, R.L., et al. (2010). Genetic Analysis of Transforming Events That Convert Chronic Myeloproliferative Neoplasms to Leukemias. Cancer Research *70*, 447–452.

Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., et al. (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood *114*, 144–147.

Adès, L., Itzykson, R., and Fenaux, P. (2014). Myelodysplastic syndromes. The Lancet.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L.A.M., et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell *121*, 295–306.

Albitar, M., Manshouri, T., Shen, Y., Liu, D., and Beran, M. (2002). Myelodysplastic syndrome is not merely "preleukemia." Blood.

Alhan, C., Westers, T.M., van der Helm, L.H., Eeltink, C., Huls, G., Witte, B.I., Buchi, F., Santini, V., Ossenkoppele, G.J., and van de Loosdrecht, A.A. (2014). Absence of aberrant myeloid progenitors by flow cytometry is associated with favorable response to azacitidine in higher risk myelodysplastic syndromes. Cytometry *86*, 207–215.

Aran, D., Sabato, S., and Hellman, A. (2013). DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biology *14*, R21.

Atak, Z.K., De Keersmaecker, K., Gianfelici, V., and Geerdens, E. (2012). High accuracy mutation detection in leukemia on a selected panel of cancer genes. PLoS ONE.

Aul, C., Bowen, D.T., and Yoshida, Y. (1998). Pathogenesis, etiology and epidemiology of myelodysplastic syndromes. Haematologica *83*, 71–86.

Álvarez-Errico, D., Vento-Tormo, R., Sieweke, M., and Ballestar, E. (2015). Epigenetic control of myeloid cell differentiation, identity and function. Nat Rev Immunol *15*, 7–17.

Balaian, L., Burkart, M., Yost, S., Rozenhak, S., and Ball, E.D. (2013a). Splicing inhibitors reduce human AML CD34+ cell survival and self-renewal during MDS/AML evolution in a leukemia stem cell supportive niche assay. Cancer Research.

Balaian, L., Crews, L.A., Zipeto, M., Anna, K., and Ball, E.D. (2013b). A Highly Selective SF3B1-Targeted Splicing Inhibitor Reduces Human CD34+ Cell Survival and Self-Renewal In Acute Myeloid Leukemia. Blood.

Bally, C., Adès, L., Renneville, A., Sebert, M., Eclache, V., Preudhomme, C., Mozziconacci, M.-J., de The, H., Lehmann-Che, J., and Fenaux, P. (2014). Leukemia Research. Leukemia Research *38*, 751–755.

Bashtrykov, P., Jankevicius, G., Smarandache, A., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2012). Specificity of Dnmt1 for Methylation of Hemimethylated CpG Sites Resides in Its Catalytic Domain. Chemistry & Biology *19*, 572–578.

Beerman, I., Bock, C., Garrison, B.S., Smith, Z.D., and Gu, H. (2013). Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. Cell Stem Cell.

Bejar, R., Lord, A., Stevenson, K., and Bar-Natan, M. (2014). TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients. Blood.

Bejar, R., Stevenson, K.E., Caughey, B.A., Abdel-Wahab, O., Steensma, D.P., Galili, N., Raza, A., Kantarjian, H., Levine, R.L., Neuberg, D., et al. (2012). Validation of a Prognostic Model and the Impact of Mutations in Patients With Lower-Risk Myelodysplastic Syndromes. Journal of Clinical Oncology *30*, 3376–3382.

Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., Levine, R.L., Neuberg, D., et al. (2011). Clinical Effect of Point Mutations in Myelodysplastic Syndromes. N. Engl. J. Med. *364*, 2496–2506.

Bellizzi, D., D'Aquila, P., Scafone, T., Giordano, M., Riso, V., Riccio, A., and Passarino, G. (2013). The Control Region of Mitochondrial DNA Shows an Unusual CpG and Non-CpG Methylation Pattern. DNA Research *20*, 537–547.

Bock, C., Beerman, I., Lien, W.H., Smith, Z.D., Gu, H., and Boyle, P. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol. Cell.

Bock, C., Tomazou, E.M., Brinkman, A.B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H.G., and Meissner, A. (2010). Quantitative comparison of genome-wide Dna methylation mapping technologies. Nat. Biotechnol. *28*, 1106–1114.

Bocker, M.T., Tuorto, F., Raddatz, G., Musch, T., Yang, F.-C., Xu, M., Lyko, F., and Breiling, A. (2012). Hydroxylation of 5-methylcytosine by TET2 maintains the active state of the mammalian HOXA cluster. Nat Commun *3*, 818.

Bonnal, S., Vigevani, L., and Valcárcel, J. (2012). The spliceosome as a target of novel antitumour drugs. Nat Rev Drug Discov *11*, 847–859.

Breccia, M., Loglisci, G., Cannella, L., Finsinger, P., Mancini, M., Serrao, A., Santopietro, M., Salaroli, A., and Alimena, G. (2012). Application of French prognostic score to patients with International Prognostic Scoring System intermediate-2 or high risk myelodysplastic syndromes treated with 5-azacitidine is able to predict overall survival and rate of response. Leuk Lymphoma *53*, 985–986.

Brooks, A.N., Yang, L., Duff, M.O., Hansen, K.D., Park, J.W., Dudoit, S., Brenner, S.E., and Graveley, B.R. (2011). Conservation of an RNA regulatory map between Drosophila and mammals. Genome Research *21*, 193–202.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Cooper, T.A. (2005). Use of minigene systems to dissect alternative splicing elements. Methods *37*, 331–340.

Couronné, L., Lippert, E., Andrieux, J., Kosmider, O., Radford-Weiss, I., Penther, D., Dastugue, N., Mugneret, F., Lafage, M., Gachard, N., et al. (2010). Analyses of TET2 mutations in post-myeloproliferative neoplasm acute myeloid leukemias. Leukemia *24*, 201–203.

Delhommeau, F., Dupont, S., Valle, Della, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.-P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in myeloid cancers. N. Engl. J. Med. *360*, 2289–2301.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., ndez, J.M.E., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. The EMBO Journal *32*, 645–655.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*, 491–498.

Douglas, A.G.L., and Wood, M.J.A. (2011). RNA splicing: disease and therapy. Brief Funct Genomics *10*, 151–164.

Ebert, B.L., Pretz, J., Bosco, J., Chang, C.Y., Tamayo, P., Galili, N., Raza, A., Root, D.E., Attar, E., Ellis, S.R., et al. (2008). Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. Nature *451*, 335–339.

Ehrlich, M., Gama-Sosa, M.A., and Huang, L.H. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids ….

Fenaux, P., and Adès, L. (2009). Leukemia Research. Leukemia Research *33*, S7–S11.

Fenaux, P., Mufti, G.J., Hellström-Lindberg, E., Santini, V., Finelli, C., Giagounidis, A., Schoch, R., Gattermann, N., Sanz, G., List, A., et al. (2009). ArticlesEfficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncology *10*, 223–232.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. Cancer Cell *18*, 553–567.

Flores, K., Wolschin, F., Corneveaux, J.J., Allen, A.N., Huentelman, M.J., and Amdam, G.V. (2012). Genome-wide association between DNAmethylation and alternative splicing in aninvertebrate. BMC Genomics *13*, 1–1.

Garcia-Manero, G. (2008). Demethylating agents in myeloid malignancies. Current Opinion in Oncology *20*, 705–710.

Garcia-Manero, G., and Fenaux, P. (2011). Hypomethylating agents and other novel strategies in myelodysplastic syndromes. Journal of Clinical Oncology *29*, 516–523.

Gelfman, S., Cohen, N., Yearim, A., and Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. Genome Research *23*, 789–799.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and EpigeneticDynamics during Specification ofHuman Embryonic Stem Cells. Cell *153*, 1149–1163.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE *5*, e15367.

Gu, H., Bock, C., Mikkelsen, T.S., Jäger, N., Smith, Z.D., Tomazou, E., Gnirke, A., Lander, E.S., and Meissner, A. (2010). Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. *7*, 133–136.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Bin Xie, Zhong, C., Hu, S., Le, T., Fan, G., et al. (2013). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nature Publishing Group *17*, 215–222.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., et al. (2013). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia *28*, 241–247.

Hanssens, K., Brenet, F., Agopian, J., Georgin-Lavialle, S., Damaj, G., Cabaret, L., Chandesris, M.O., de Sepulveda, P., Hermine, O., Dubreuil, P., et al. (2014). SRSF2-p95 hotspot mutation is highly associated with advanced forms of mastocytosis and mutations in epigenetic regulator genes. Haematologica *99*, 830–835.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science *333*, 1303–1307.

Hofmann, W.-K., and Koeffler, H.P. (2005). Myelodysplastic syndrome. Annu. Rev. Med. *56*, 1–16.

Hsieh, C.L. (2000). Dynamics of DNA methylation pattern. Current Opinion in Genetics & Development.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genet *6*, e1001134.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science *333*, 1300–1303.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Itzykson, R., Kosmider, O., Cluzeau, T., Mas, V.M.-D., Dreyfus, F., Beyne-Rauzy, O., Quesnel, B., Vey, N., Gelsi-Boyer, V., Raynaud, S., et al. (2011a). leu201171a. Leukemia *25*, 1147–1152.

Itzykson, R., Thépot, S., Quesnel, B., Dreyfus, F., Beyne-Rauzy, O., Turlure, P., Vey, N., Recher, C., Dartigeas, C., Legros, L., et al. (2011b). Prognostic factors for response and overall survival in 282 patients with higher-risk myelodysplastic syndromes treated with azacitidine. Blood *117*, 403–411.

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.-A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biology *14*, R119.

Iwasaki, H., and Akashi, K. (2007). Hematopoietic developmental pathways: on cellular basis. Oncogene *26*, 6687–6696.

Jadersten, M., Saft, L., Pellagatti, A., Gohring, G., Wainscoat, J.S., Boultwood, J., Porwit, A., Schlegelberger, B., and Hellstrom-Lindberg, E. (2009). Clonal heterogeneity in the 5q- syndrome: p53 expressing progenitors prevail during lenalidomide treatment and expand at disease progression. Haematologica *94*, 1762–1766.

Jankowska, A.M., Szpurka, H., Tiu, R.V., Makishima, H., Afable, M., Huh, J., O'Keefe, C.L., Ganetzky, R., McDevitt, M.A., and Maciejewski, J.P. (2009). Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. Blood *113*, 6403–6410.

Jiang, Y., Dunbar, A., Gondek, L.P., Mohan, S., Rataul, M., O'Keefe, C., Sekeres, M., Saunthararajah, Y., and Maciejewski, J.P. (2009). Aberrant DNA methylation is a

dominant mechanism in MDS progression to AML. Blood *113*, 1315–1325.

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al. (2012). KhareNatStructMolBio2012. Nature Structural & Molecular Biology *19*, 1037–1043.

Kiel, M.J., Yilmaz, O.H., Iwashita, T., Yilmaz, O.H., Terhorst, C., and Morrison, S.J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell *121*, 1109–1121.

Kinney, S.M., Chin, H.G., Vaisvila, R., Bitinaite, J., Zheng, Y., Estève, P.-O., Feng, S., Stroud, H., Jacobsen, S.E., and Pradhan, S. (2011). Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. J. Biol. Chem. *286*, 24685–24693.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in Biochemical Sciences *31*, 89–97.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839–843.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.

Kroenke, J., Fink, E., Hurst, S., Udeshi, N.D., Svinkina, T., Schneider-Kramann, R., McConkey, M.E., Jaraas, M., Bullinger, L., Carr, S.A., et al. Lenalidomide Induces Ubiquitination and Degradation of CSNK1A1 in MDS with Del(5q). Nature.

Kunimoto, H., Fukuchi, Y., Sakurai, M., Sadahira, K., Ikeda, Y., Okamoto, S., and Nakajima, H. (2012). Tet2 disruption leads to enhanced self-renewal and altered differentiation of fetal liver hematopoietic stem cells. Sci. Rep. *2*.

Langemeijer, S.M.C., Kuiper, R.P., Berends, M., Knops, R., Aslanyan, M.G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A.G., Raymakers, R.A.P., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet *41*, 838–842.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509–4518.

List, A., Dewald, G., Bennett, J., Giagounidis, A., Raza, A., Feldman, E., Powell, B., Greenberg, P., Thomas, D., and Stone, R. (2006). Lenalidomide in the myelodysplastic syndrome with chromosome 5q deletion. New England Journal of Medicine *355*, 1456–

1465.

Liu, X., Zhang, Q., Zhang, D.-E., Zhou, C., Xing, H., Tian, Z., Rao, Q., Wang, M., and Wang, J. (2009). Overexpression of an isoform of AML1 in acute leukemia and its potential role in leukemogenesis. Leukemia *23*, 739–745.

Lorenz, E., Uphoff, D., Reid, T.R., and Shelton, E. (1951). Modification of Irradiation Injury in Mice and Guinea Pigs by Bone Marrow Injections. J Natl Cancer Inst *12*, 197–201.

Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia *17*, 637–641.

Mahfouz, R.Z., Jankowska, A., Ebrahem, Q., Gu, X., Visconte, V., Tabarroki, A., Terse, P., Covey, J., Chan, K., Ling, Y., et al. (2013). Increased CDA Expression/Activity in Males Contributes to Decreased Cytidine Analog Half-Life and Likely Contributes to Worse Outcomes with 5-Azacytidine or Decitabine Therapy. Clinical Cancer Research *19*, 938–948.

Manev, H., Dzitoyeva, S., and Chen, H. (2012). Mitochondrial DNA: A Blind Spot in Neuroepigenetics. Biomolecular Concepts *3*.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. *23*, 1256–1269.

Maximow, A.A. (1909). The Lymphocyte as a stem cell common to different blood elements in embryonic development and during the post fetal life of mammals. Folia Haematologica *8*, 125–134.

Meissner, A. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Research *33*, 5868–5877.

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell.

Mittler, G., Butter, F., and Mann, M. (2009). A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Research *19*, 284–293.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 Loss Leads to Increased HematopoieticStem Cell Self-Renewal and Myeloid Transformation. Cancer Cell *20*, 11–24.

Mullighan, C.G. (2009). TET2 mutations in myelodysplasia and myeloid malignancies.

Nat Genet *41*, 766–767.

Murphy, D.M., Bejar, R., Stevenson, K., Neuberg, D., Shi, Y., Cubrich, C., Richardson, K., Eastlake, P., Garcia-Manero, G., Kantarjian, H., et al. (2013). NRAS mutations with low allele burden have independent prognostic significance for patients with lower risk myelodysplastic syndromes. Leukemia *27*, 2077–2081.

Nestor, C.E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J.M., Harrison, D.J., and Meehan, R.R. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Research *22*, 467–477.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. Cell *144*, 296–309.

Ooi, S.K.T., and Bestor, T.H. (2008). The Colorful History of Active DNA Demethylation. Cell *133*, 1145–1148.

Papaemanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616–3617.

Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat. Rev. Mol. Cell Biol. *14*, 341–356.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brodno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *IN PRESS*, 1–5.

Pellagatti, A., and Boultwood, J. (2015). The molecular pathogenesis of the myelodysplastic syndromes. Eur. J. Haematol.

Qin, T., Castoro, R., Ahdab, El, S., Jelinek, J., Wang, X., Si, J., Shu, J., He, R., Zhang, N., Chung, W., et al. (2011). Mechanisms of Resistance to Decitabine in the Myelodysplastic Syndrome. PLoS ONE *6*, e23372.

Quivoron, C., Couronné, L., Valle, Della, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.-H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell *20*, 25–38.

Rollison, D.E., Howlader, N., Smith, M.T., Strom, S.S., Merritt, W.D., Ries, L.A., Edwards, B.K., and List, A.F. (2008). Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. Blood *112*, 45–52.

Rothrock, C., Cannon, B., Hahm, B., and Lynch, K.W. (2003). A conserved signal-responsive sequence mediates activation-induced alternative splicing of CD45. Mol. Cell

12, 1317–1324.

Saint-Martin, C., Leroy, G., Delhommeau, F., Panelatti, G., Dupont, S., James, C., Plo, I., Bordessoule, D., Chomienne, C., Delannoy, A., et al. (2009). Analysis of the ten-eleven translocation 2 (TET2) gene in familial myeloproliferative neoplasms. Blood *114*, 1628–1632.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). correspondence. Nat Meth *11*, 783–784.

Saunthararajah, Y. (2013). Key clinical observations after 5-azacytidine and decitabine treatment of myelodysplastic syndromes suggest practical solutions for better outcomes. Hematology Am Soc Hematol Educ Program *2013*, 511–521.

Schübeler, D. (2015). Function and information content of DNA methylation. Nature.

Scourzic, L., Mouly, E., and Bernard, O.A. (2015). TET proteins and the control of cytosine demethylation in cancer. Genome Med *7*, 9.

Sekeres, M.A. (2011). Lenalidomide in MDS: 4th time's a charm. Blood *118*, 3757–3758.

Shafi, R., Iyer, S.P., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., and Marth, J.D. (2000). The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. Proc. Natl. Acad. Sci. U.S.a. *97*, 5735–5739.

Shen, L., Kantarjian, H., Guo, Y., Lin, E., Shan, J., Huang, X., Berry, D., Ahmed, S., Zhu, W., Pierce, S., et al. (2010). DNA Methylation Predicts Survival and Response to Therapy in Patients With Myelodysplastic Syndromes. Journal of Clinical Oncology *28*, 605–613.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79.

Singh, G., and Cooper, T.A. (2006). Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing.

Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. *29*, 68–72.

Steensma, D.P., Baer, M.R., Slack, J.L., Buckstein, R., Godley, L.A., Garcia-Manero, G., Albitar, M., Larsen, J.S., Arora, S., Cullen, M.T., et al. (2009). Multicenter Study of Decitabine Administered Daily for 5 Days Every 4 Weeks to Adults With Myelodysplastic Syndromes: The Alternative Dosing for Outpatient Treatment (ADOPT) Trial. Journal of Clinical Oncology *27*, 3842–3848.

Stone, R.M. (2009). How I treat patients with myelodysplastic syndromes. Blood *113*,

6296–6303.

Stresemann, C., and Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int. J. Cancer *123*, 8–13.

Swerdllow, S.H., Campo, E., and Harris, N.L. (2008). WHO classification of tumours of haematopoietic and lymphoid tissues.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930–935.

Tefferi, A., Lim, K.-H., Abdel-Wahab, O., Lasho, T.L., Patel, J., Patnaik, M.M., Hanson, C.A., Pardanani, A., Gilliland, D.G., and Levine, R.L. (2009a). Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. Leukemia *23*, 1343–1345.

Tefferi, A., Pardanani, A., Lim, K.-H., Abdel-Wahab, O., Lasho, T.L., Patel, J., Gangat, N., Finke, C.M., Schwager, S., Mullally, A., et al. (2009b). TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. Leukemia *23*, 905–911.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442–449.

Tsagaratou, A., Äijö, T., Lio, C.-W.J., Yue, X., Huang, Y., Jacobsen, S.E., Lähdesmäki, H., and Rao, A. (2014). Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proc. Natl. Acad. Sci. U.S.a. *111*, E3306–E3315.

Valencia, A., Masala, E., Rossi, A., Martino, A., Sanna, A., Buchi, F., Canzian, F., Cilloni, D., Gaidano, V., Voso, M.T., et al. (2014). Expression of nucleoside-metabolizing enzymes inmyelodysplastic syndromes and modulation of responseto azacitidine. *28*, 621–628.

van der Ploeg, R. (2009). Stable-isotope labeling by amino acids in cell culture (SILAC).

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., et al. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol. Cell *49*, 645–656.

Walter, M.J., Shen, D., Shao, J., Ding, L., White, B.S., Kandoth, C., Miller, C.A., Niu, B., McLellan, M.D., Dees, N.D., et al. (2013). Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia *27*, 1275–1282.

Walter, M.J., Shen, D., Ding, L., Shao, J., Koboldt, D.C., Chen, K., Larson, D.E.,

McLellan, M.D., Dooling, D., Abbott, R., et al. (2012). Clonal Architecture of Secondary Acute Myeloid Leukemia. N. Engl. J. Med. *366*, 1090–1098.

Will, B., Zhou, L., Vogler, T.O., Ben-Neriah, S., Schinke, C., Tamari, R., Yu, Y., Bhagat, T.D., Bhattacharyya, S., Barreyro, L., et al. (2012). Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations. Blood *120*, 2076–2086.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Woll, P.S., Kjällquist, U., Chowdhury, O., Doolittle, H., Wedge, D.C., Thongjuea, S., Erlandsson, R., Ngara, M., Anderson, K., Deng, Q., et al. (2014). Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. Cancer Cell *25*, 794–808.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389–393.

Yamazaki, J., Taby, R., Vasanthakumar, A., Macrae, T., Ostler, K.R., Shen, L., Kantarjian, H.M., Estecio, M.R., Jelinek, J., Godley, L.A., et al. (2014). Effects of TET2mutations on DNA methylation in chronic myelomonocytic leukemia. Epigenetics *7*, 201–207.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., and Ee, L.S. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell.

Yilmaz, O.H., Kiel, M.J., and Morrison, S.J. (2006). SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. Blood *107*, 924–930.

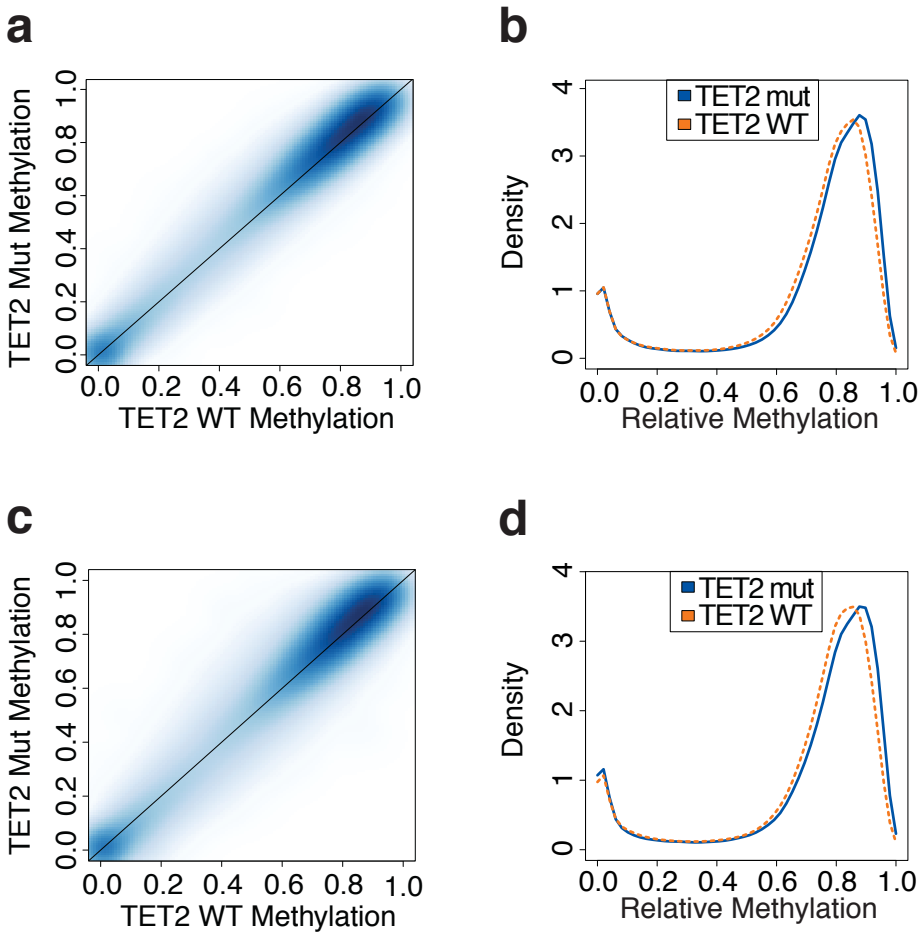
Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69.

Zeidan, A.M., Lee, J.-W., Prebet, T., Greenberg, P., Sun, Z., Juckett, M., Smith, M.R., Paietta, E., Gabrilove, J., Erba, H.P., et al. (2014). Comparison of the prognostic utility of the revised International Prognostic Scoring System and the French Prognostic Scoring System in azacitidine-treated patients with myelodysplastic syndromes. Br J Haematol *166*, 352–359.

Appendix I

Supplemental Table 3.1: Somatic mutations in individual MDS samples.

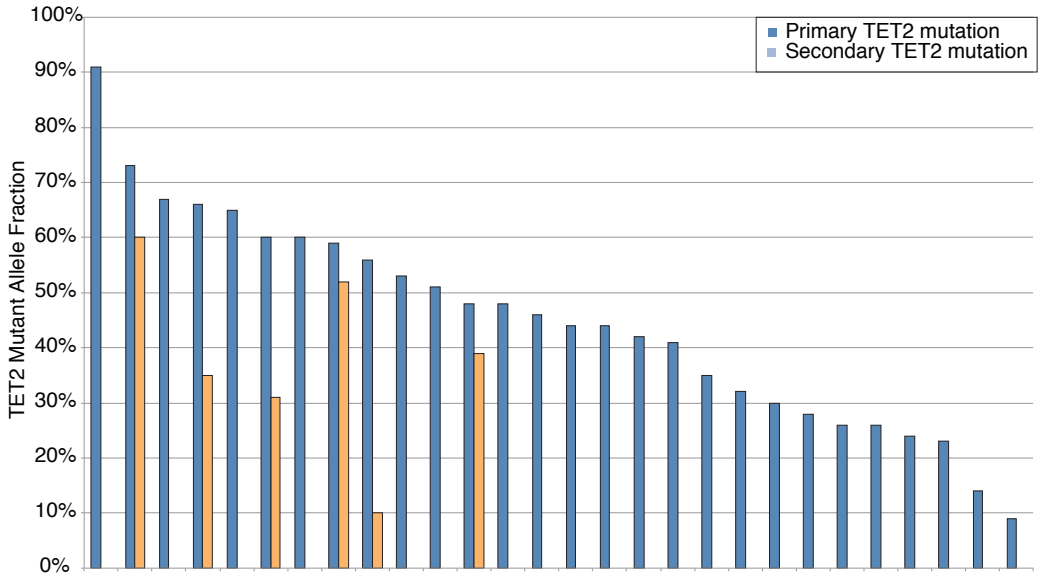
TET2 mutation	Other known somatic mutations
(c.3594+5G>A)	
(c.3955-2A>G)	SF3B1 p.K666T
p.C1135Y	SRSF2 p.P95L
p.C1193W	
p.E1082AfsX21 p.N275IfsX18	SF3B1 p.E622D
p.E1165K p.P761LfsX52	SF3B1 p.H662Y
p.E798*	EZH2 p.D233G, SF3B1 p.K700E
p.G1861V	SRSF2 p.P95_R102del
p.L1790TfsX32	TP53 p.Q136*
p.N374KfsX3	SRSF2 p.P95H
p.Q243* p.N482IfsX4	SF3B1 p.K700E
p.Q599* p.Y1589*	SF3B1 p.H662Y
p.Q745*	
p.S1583* p.A1768V p.R1176G	
TET2 p.L920FfsX4	DNMT3a pR736H
TET2 p.N275IfsX18 p.P1594QfsX16	DNMT3a p.R882H
TET2_C1135Y	
TET2_E184AfsX7	SRSF2 P95L
TET2_E852*	SF3B1 p.K700E
TET2_K1339*	
TET2_P1725T	JAK2-V617F, ASXL1-P920TfsX4, U2AF1 p.Q157P
TET2_Q1527*	SRSF2 p.P95H
TET2_Q1537*	SF3B1 K700E
TET2_Q531*	DNMT3a p.W581R, SF3B1 p.PK700E
TET2_Q642* TET2_N1266S	RUNX1 p.H404PfsX196 ASXL1 p.E635RfsX15 EZH2 p.R298H p.R495GfsX19
TET2_Q685*	
TET2_Q884*	
TET2_R1261C	
TET2_R1261H	NRAS-G12S NRAS-G13D
TET2_R1465*	SRSF2 P95R
TET2_R1516* TET2_T1884A	ASXL1 p.Y591*
TET2_R544* TET2_M1333K	ASXL1 p.E635RfsX15 EZH2 p.I131_P132delinsMT
TET2_S1583* TET2_H1817N TET2_A1768V	
TET2_S643YfsX35 TET2_F1300*	
	SRSF2 P95R
	p53 E336GfsX9
	NRAS-G12S NRAS-G13D, ASXL1 p.L697RfsX14, SRSF2 P95H
	ASXL1 p.S903*
	ASXL1 p.E635RfsX15
	JAK2-V617F, ASXL1-Y591*, DNMT3a p.W409*
	SF3B1 p.K666N, SRSF2 p.95H
	DNMT3a p.G543D, SF3B1 p.K700E
	SF3B1 p.N626D, SRSF2 p.P95H
	DNMT3a p.R736C, SF3B1 p.K700E
	DNMT3a (c.1667+1G>T), SF3B1 p.K700E
	RUNX1 p.M133_A134insV EZH2 p.C457Y (c.2111-2A>T)
	ASXL1 p.Q757*, DNMT3a p.R635W, SF3B1 p.R625C
	SRSF2 p.P95H
	TP53 p.A159P
	EZH2 p.Y663H, SF3B1 p.K700E
	SF3B1 p.K700E
	SRSF2 p.P95H
	SRSF2 p.P95_R102del
	SF3B1 p.H662Y
	SF3B1 p.K700E



Supplemental Figure 3.1: Global methylation in MDS Samples. All samples (a, b) and high allele fraction samples (c, d). Comparison of average methylation between TET2 mutant and TET2 wildtype groups, where darker blue indicates increased density (a, c) and relative methylation of each group as a function of CpG density (b, d).

Supplemental Table 3.2: Characteristics of high allele fraction MDS patient samples. Statistics: p-values calculated using two-tailed independent t-test. Unless indicated, p = n.s.

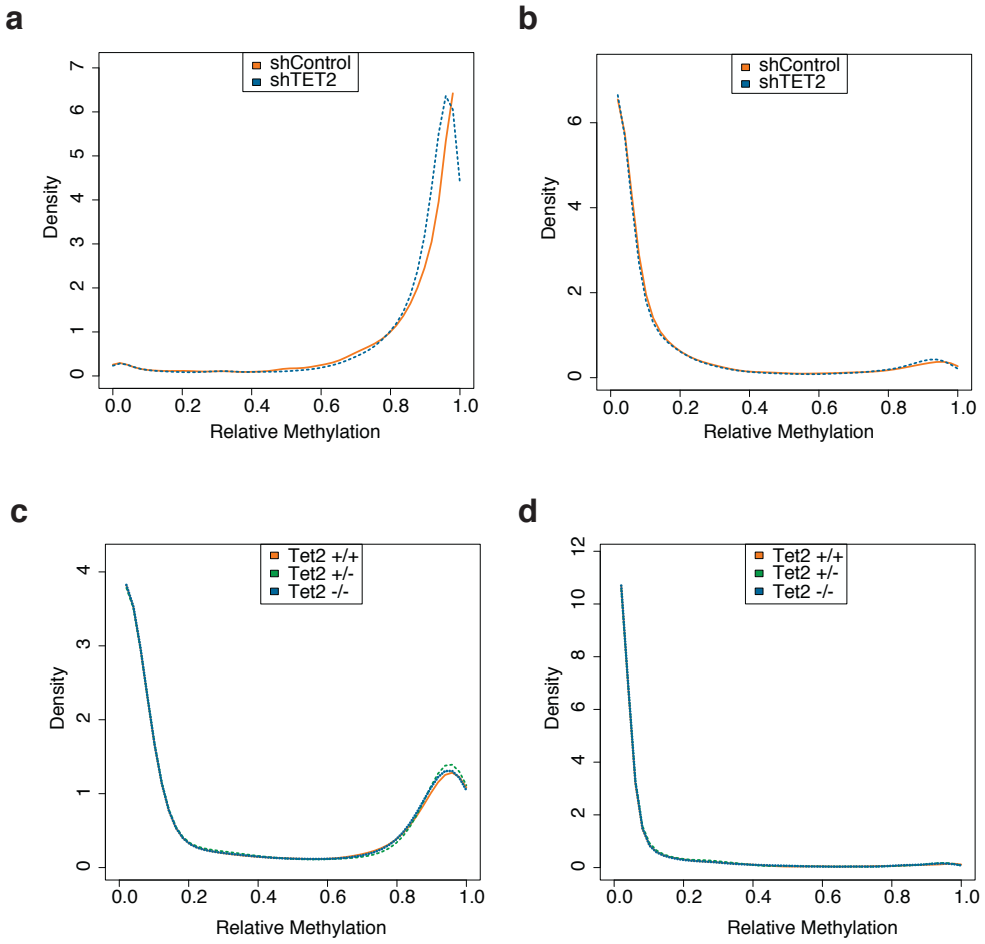
	TET2 WT	TET2 Mutant	
n =	17	17	
Age (SD)	70.06 (6.58)	70.71 (8.51)	
Sex			
M	13	17	
F	4	0	p=0.04*
FAB			
RA	7	7	
RARS	4	3	
RAEB/RAEB-t	6	7	
AML	0	0	
IPSS			
Low	4	5	
Int1	8	8	
Int2	4	4	
Unknown	1	0	
CBC			
% Blast (SD)	5.06 (4.78)	4.24 (4.18)	
Cellularity (SD)	64.38 (24.35)	66.77 (30.41)	
Hgb (SD)	9.58 (1.55)	10.15 (1.46)	
WBC (SD)	3.97 (2.46)	7.86 (16.76)	
ANC (SD)	16.66 (25.14)	8.79 (14.56)	
Lymph (SD)	36.91 (24.90)	43.27 (19.98)	
Mono (SD)	11.08 (9.45)	8.88 (6.05)	
Platelets (SD)	169.24 (164.57)	125.59 (129.08)	
Other Mutations			
RUNX1	1	1	
TP53	1	1	
JAK2	1	1	
NRAS	0	0	
ASXL1	4	4	
EZH2	2	3	
DNMT3A	3	1	
SF3B1	5	3	
SRSF2	3	5	
U2AF1	0	1	



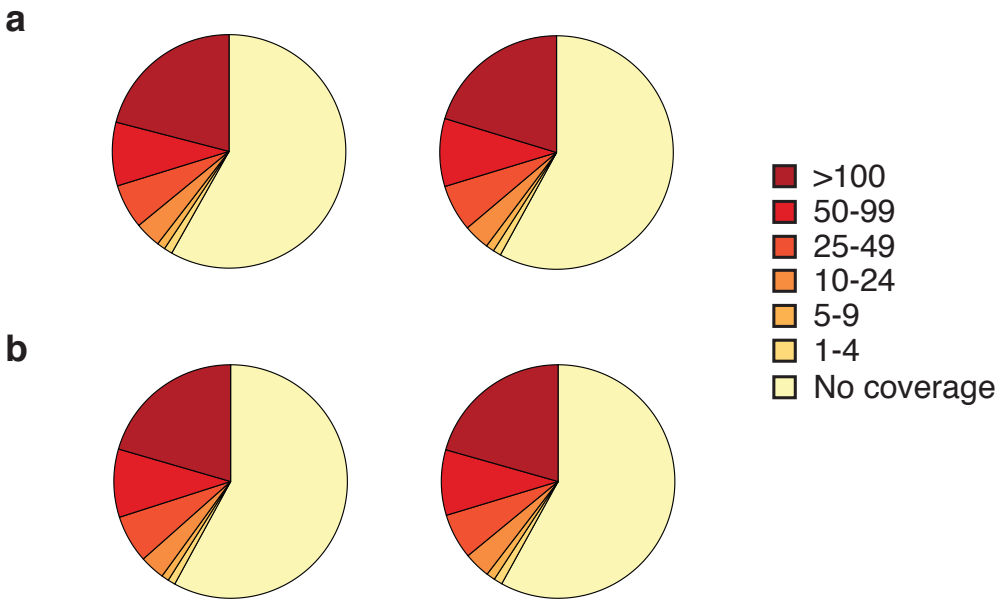
Supplemental Figure 3.2: TET2 VAF in TET2-mutant MDS samples. Variant allele fractions could be estimated based on MALDI-TOF mass spectrometry (Sequenom) for the majority of MDS samples.

Supplemental Table 3.3: Methylation and CpG Content for all genomic regions.

Region	Hypermethylated CpGs (% Total Hypermethylated)	Hypomethylated CpGs (% Total Hypomethylated)	CpGs (% Total CpGs)
Promoter	94 (5.7)	46 (15.5)	230289 (35.0)
CGI	148 (9.0)	106 (35.7)	296317 (45.0)
CGI Shores	260 (15.8)	105 (35.4)	110867 (16.8)
Non-Promoter Intron-Exon Junctions	351 (21.4)	54 (18.2)	77824 (11.8)
Enhancer	90 (5.5)	39 (13.1)	161912 (24.6)
LINE	69 (4.2)	3 (1.0)	13550 (2.1)
SINE	160 (9.7)	25 (8.4)	120436 (18.3)
LTR	107 (6.5)	8 (2.7)	19949 (3.0)
3' UTR	44 (2.7)	13 (4.4)	10066 (1.5)
5' UTR	107 (6.5)	25 (8.4)	109844 (16.7)
Introns	842 (51.2)	142 (47.8)	315697 (47.9)
Exons	269 (16.4)	106 (35.7)	146583 (22.3)
Introns (Junctions Excluded)	736 (44.8)	116 (39.1)	214834 (32.6)
Exons Junctions Excluded)	44 (2.7)	13 (4.4)	17313 (2.6)
None	319 (19.4)	39 (13.1)	60499 (9.2)
All	1644 (100)	297 (100)	658609 (100)



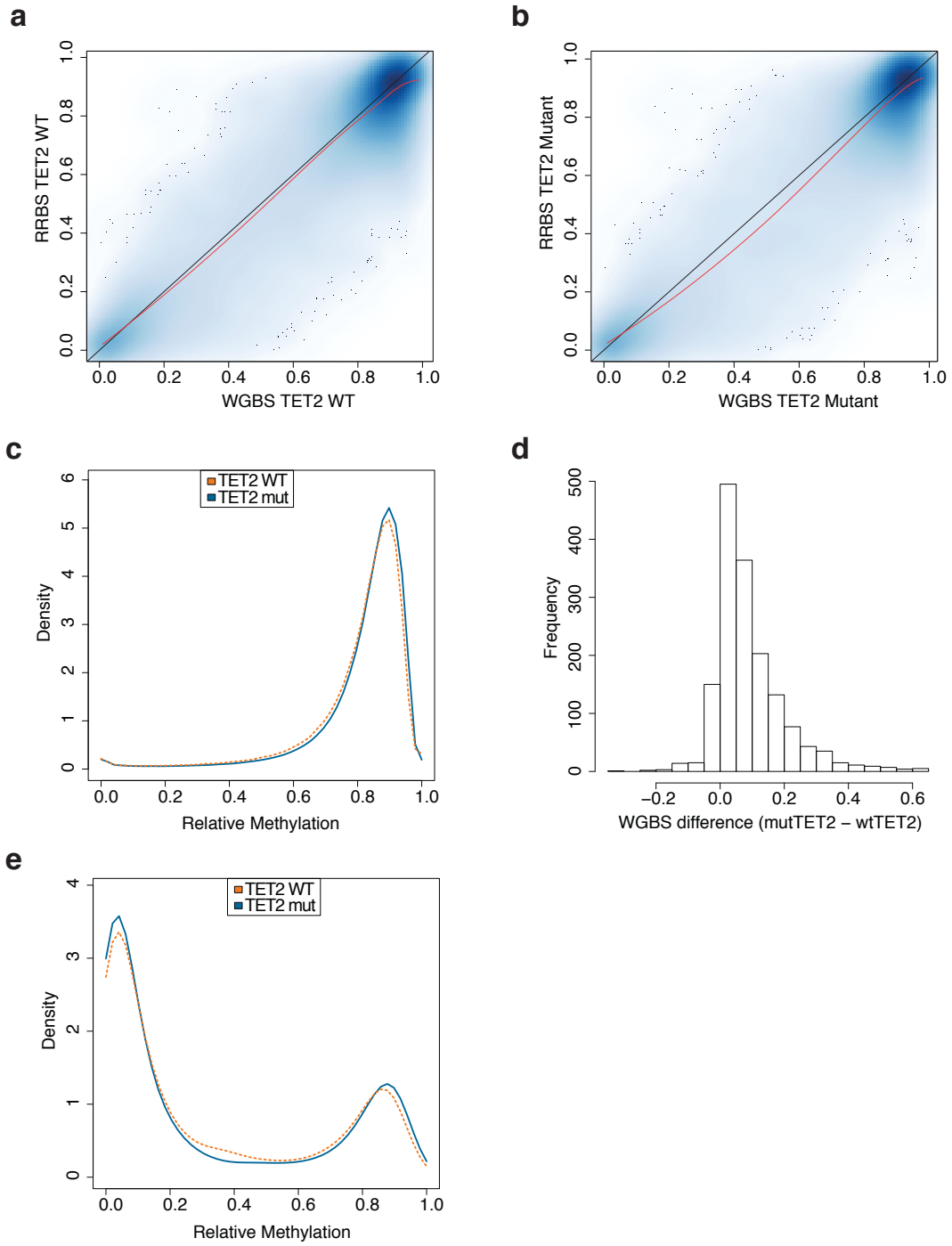
Supplemental Figure 3.3: Intron-Exon and promoter methylation in *TET2/Tet2* deficient CD34+ cells and mouse LSK cells. Relative methylation is shown as a function of density in CD34+ cells transduced with shRNA against *TET2* or control (a, b) and sorted murine LSK cells from pI:pC-treated mice (c, d). All mice are Mx-Cre^{+}, genotypes as indicated. Regions shown are intron-exon boundaries excluding promoters (a), intron-exon boundaries including promoters (c), and promoters (b, d).



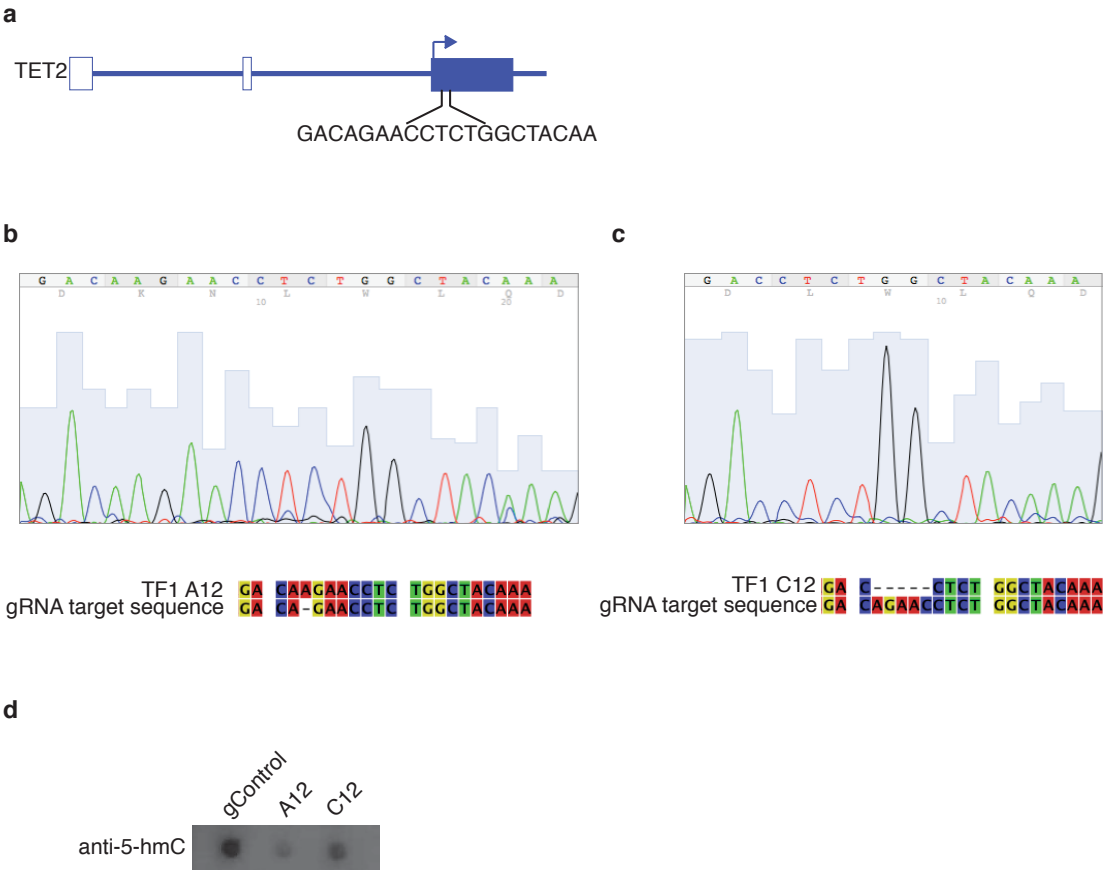
Supplemental Figure 3.4: Representative sequencing coverage (RRBS). Number of global reads are shown for two representative *TET2*-wildtype (a) and *TET2*-mutant (b) MDS samples.

Supplemental Table 3.4: Characteristics of MDS samples used for WGBS.

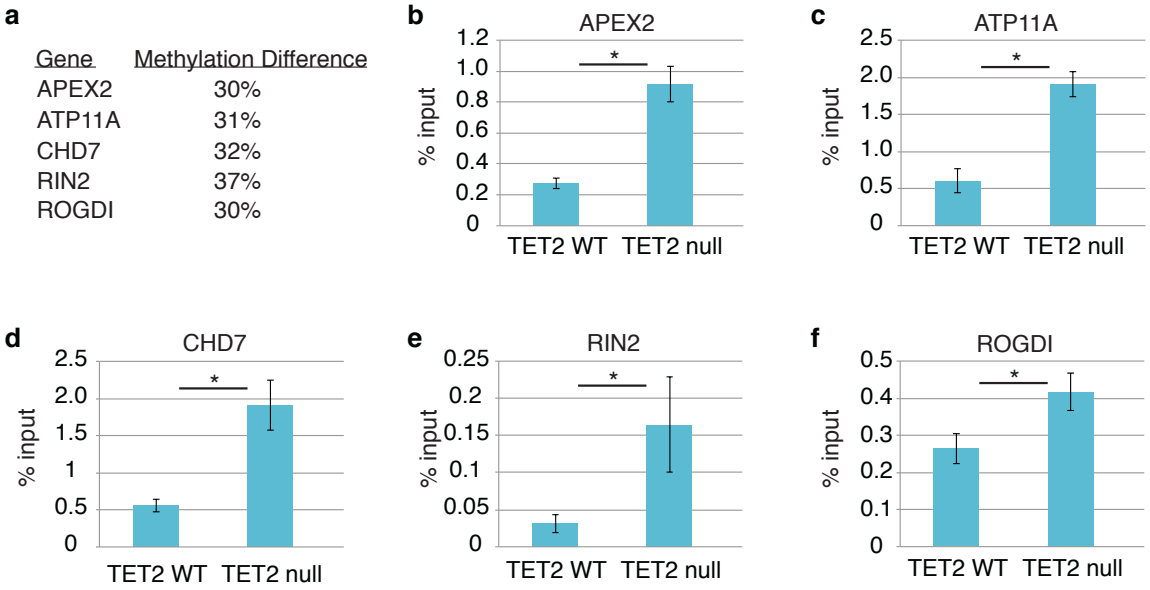
	TET2 WT	TET2 Mutant
n =	2	2
Age (SD)	62.5 (12.02)	60 (8.49)
Sex		
M	0	2
F	2	0
FAB		
RA	1	1
RARS	1	1
RAEB/RAEB-t	0	0
AML	0	0
IPSS		
Low	2	2
CBC		
% Blast (SD)	2 (1.41)	2 (1.41)
Cellularity (SD)	70 (0)	67.5 (38.89)
HGB (SD)	8.55 (2.05)	11.1 (0.85)
WBC (SD)	8.58 (0.32)	3.49 (1.58)
ANC (SD)	5.92 (0.34)	1.35 (0.83)
Lymph (SD)	21.5 (4.95)	51.5 (2.12)
Mono (SD)	6.5 (3.54)	10 (7.07)
Platelets (SD)	246 (151.32)	216.5 (20.51)
Other Mutations		
RUNX1	0	0
TP53	0	0
JAK2	0	0
NRAS	0	0
ASXL1	0	1
EZH2	0	0
DNMT3A	0	1
SF3B1	1	1
SRSF2	1	0
U2AF1	0	0



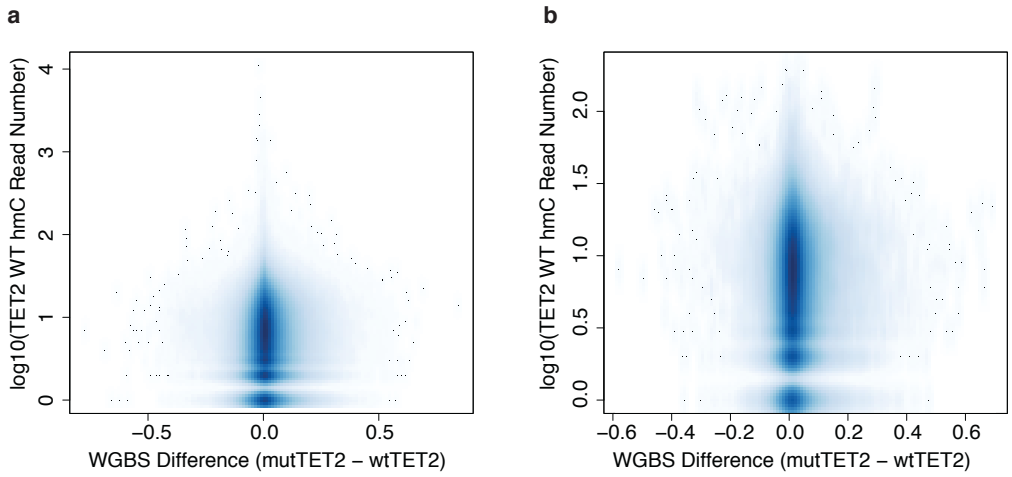
Supplemental Figure 3.5: Global methylation and promoter methylation in MDS samples by WGBS. Comparison of methylation by RRBS (y-axis) and WGBS (x-axis in *TET2*-wildtype samples (a) and TET2-mutant samples (b). Red line indicates best-fit by LOESS. Global methylation (c) and promoter methylation (e) by WGBS is shown as a function of density (c) and global differential methylation (relative *TET2*-mutant methylation—relative *TET2*-wildtype methylation) is shown by frequency (d).



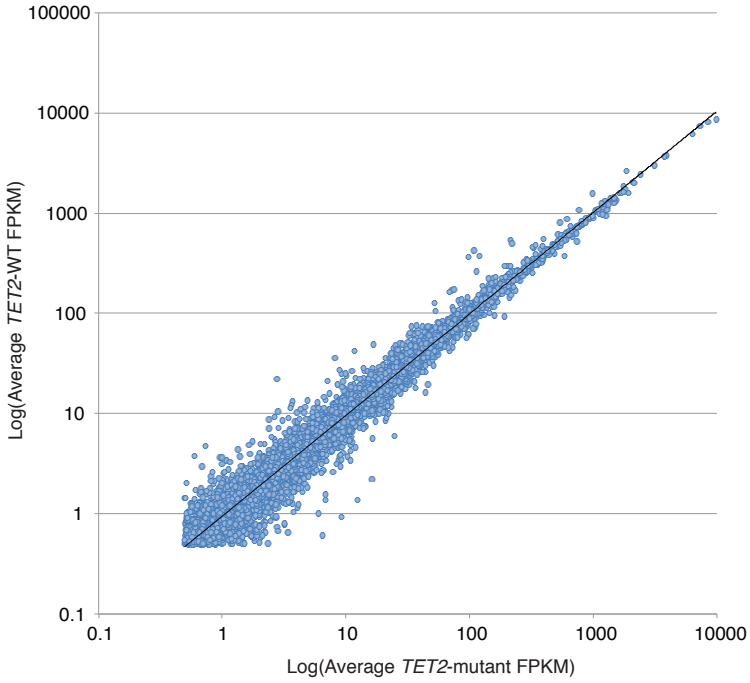
Supplemental Figure 3.6: Engineering targeted mutations in *TET2* using the CRISPR-Cas9 system in TF1 cells. The first three exons of TET2, including the guide RNA target sequence (inset) in exon 3 (a). Sequencing results for two separate single-cell-derived TF1 clones the Cas9 nuclease and gRNA targeting *TET2*. Clone A12 (b) contains a single nucleotide insertion, while clone C12 (c) contains a 5-nucleotide deletion. Dotblot for 5-hmC in genomic DNA of control, A12, and C12 TF1 cell clones (d).



Supplemental Figure 3.7: Validation of RRBS/WGBS in TF1 cells by Me-DIP-PCR. A subset of highly differentially methylated regions (*TET2*-mutant – *TET2*-wildtype) in MDS samples (a) was assessed for mC content in TF1 cells by Me-DIP-PCR (see methods for details). mC content for each region displayed as percent of total normalized to input DNA (b-f). All immunoprecipitations used standard high salt conditions, except for RIN2, ROGDI (low salt). Error bars represent SEM; significance calculated by two-tailed t-test (* indicated $p \leq 0.05$).



Supplemental Figure 3.8: Overlap of hmC and WGBS data. Regions with good coverage (>5 reads) in both RRHS (y-axis) and WGBS intron-exon (x-axis) datasets were compared. WGBS reads were plotted by normalized differential methylation (normalized methylation for *TET2*-wildtype samples were subtracted from normalized methylation values in *TET2*-mutant samples). Panel A shows all overlapping regions, while Panel B focuses on highly overlapping regions.



Supplemental Figure 3.9: Correlation analysis of RNA-Seq data. Log-transformed FPKM values were plotted for all well-covered (Log(FPKM)>5) reads in *TET2*-wildtype (y-axis) and *TET2*-mutant TF1 cell samples. Best-fit line was generated using linear regression, $R^2 = 0.97301$.

Supplemental Table 4.1: Selected target regions included coding exons of recurrently mutated genes plus an additional 10 nucleotides on either side to encompass consensus splicing sequences.

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr1	43814924	43815040	*MPL*	e10
chr1	115251142	115251281	*NRAS*	e5
chr1	115252180	115252359	*NRAS*	e4
chr1	115256411	115256609	*NRAS*	e3
chr1	115258661	115258791	*NRAS*	e2
chr2	25457138	25457299	*DNMT3A*	e23
chr2	25458566	25458704	*DNMT3A*	e22
chr2	25459795	25459884	*DNMT3A*	e21
chr2	25461989	25462094	*DNMT3A*	e20
chr2	25463161	25463329	*DNMT3A*	e19
chr2	25463499	25463609	*DNMT3A*	e18
chr2	25464421	25464586	*DNMT3A*	e17
chr2	25466757	25466861	*DNMT3A*	e16
chr2	25467014	25467217	*DNMT3A*	e15
chr2	25467399	25467531	*DNMT3A*	e14
chr2	25468112	25468211	*DNMT3A*	e13
chr2	25468879	25468943	*DNMT3A*	e12
chr2	25469019	25469188	*DNMT3A*	e11
chr2	25469479	25469655	*DNMT3A*	e10
chr2	25469910	25470037	*DNMT3A*	e9
chr2	25470450	25470628	*DNMT3A*	e8
chr2	25470896	25471131	*DNMT3A*	e7
chr2	25497800	25497966	*DNMT3A*	e6
chr2	25498359	25498422	*DNMT3A*	e5
chr2	25505300	25505590	*DNMT3A*	e4
chr2	25522998	25523122	*DNMT3A*	e3
chr2	25536772	25536863	*DNMT3A*	e2
chr2	198257017	198257195	*SF3B1*	e25
chr2	198257686	198257922	*SF3B1*	e24
chr2	198260770	198261062	*SF3B1*	e23
chr2	198262699	198262850	*SF3B1*	e22
chr2	198263175	198263315	*SF3B1*	e21
chr2	198264769	198264900	*SF3B1*	e20
chr2	198264966	198265168	*SF3B1*	e19
chr2	198265429	198265670	*SF3B1*	e18
chr2	198266114	198266259	*SF3B1*	e17
chr2	198266456	198266622	*SF3B1*	e16
chr2	198266699	198266864	*SF3B1*	e15
chr2	198267270	198267560	*SF3B1*	e14
chr2	198267663	198267769	*SF3B1*	e13
chr2	198268299	198268498	*SF3B1*	e12
chr2	198269790	198269911	*SF3B1*	e11
chr2	198269989	198270206	*SF3B1*	e10
chr2	198272712	198272853	*SF3B1*	e9
chr2	198273083	198273315	*SF3B1*	e8
chr2	198274484	198274741	*SF3B1*	e7
chr2	198281455	198281645	*SF3B1*	e6
chr2	198283223	198283322	*SF3B1*	e5
chr2	198285142	198285276	*SF3B1*	e4
chr2	198285743	198285867	*SF3B1*	e3
chr2	198288522	198288708	*SF3B1*	e2
chr2	198299686	198299733	*SF3B1*	e1

Supplemental Table 4.1, continued

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr2	209113083	209113394	*IDH1*	e4
chr3	105438881	105439104	*CBLB*	e10
chr3	105452843	105452994	*CBLB*	e9
chr3	128199852	128200171	*GATA2*	e6
chr3	128200652	128200797	*GATA2*	e5
chr3	128202693	128202858	*GATA2*	e4
chr3	128204560	128205221	*GATA2*	e3
chr3	128205636	128205884	*GATA2*	e2
chr4	55593572	55593718	*KIT*	e11
chr4	55599226	55599368	*KIT*	e17
chr4	106155090	106158518	*TET2*	e3
chr4	106162486	106162596	*TET2*	e4
chr4	106163981	106164094	*TET2*	e5
chr4	106164717	106164945	*TET2*	e6
chr4	106180766	106180936	*TET2*	e7
chr4	106182906	106183015	*TET2*	e8
chr4	106190757	106190914	*TET2*	e9
chr4	106193711	106194085	*TET2*	e10
chr4	106196195	106197686	*TET2*	e11
chr5	170834694	170834788	*NPM1*	e10
chr5	170837521	170837579	*NPM1*	e11
chr7	139044998	139045078	*LUC7L2*	e1
chr7	139060798	139060912	*LUC7L2*	e2
chr7	139083335	139083453	*LUC7L2*	e3
chr7	139086873	139087003	*LUC7L2*	e4
chr7	139090380	139090543	*LUC7L2*	e5
chr7	139091910	139092106	*LUC7L2*	e6
chr7	139094299	139094410	*LUC7L2*	e7
chr7	139097287	139097336	*LUC7L2*	e8
chr7	139102274	139102485	*LUC7L2*	e9
chr7	139106899	139107096	*LUC7L2*	e10
chr7	140453065	140453203	*BRAF*	e15
chr7	148504728	148504808	*EZH2*	e20
chr7	148506153	148506257	*EZH2*	e19
chr7	148506392	148506492	*EZH2*	e18
chr7	148507415	148507516	*EZH2*	e17
chr7	148508707	148508822	*EZH2*	e16
chr7	148511041	148511239	*EZH2*	e15
chr7	148511996	148512141	*EZH2*	e14
chr7	148512588	148512648	*EZH2*	e13
chr7	148513766	148513880	*EZH2*	e12
chr7	148514304	148514493	*EZH2*	e11
chr7	148514959	148515219	*EZH2*	e10
chr7	148516678	148516789	*EZH2*	e9
chr7	148523536	148523734	*EZH2*	e8
chr7	148524246	148524368	*EZH2*	e7
chr7	148525822	148525982	*EZH2*	e6
chr7	148526810	148526950	*EZH2*	e5
chr7	148529716	148529852	*EZH2*	e4
chr7	148543552	148543700	*EZH2*	e3
chr7	148544264	148544400	*EZH2*	e2
chr9	5073688	5073795	*JAK2*	e14
chr11	32410594	32410735	*WT1*	e10
chr11	32413508	32413620	*WT1*	e9
chr11	32414202	32414311	*WT1*	e8

Supplemental Table 4.1, continued

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr11	32417793	32417963	*WT1*	e7
chr11	32421484	32421600	*WT1*	e6
chr11	32438026	32438096	*WT1*	e5
chr11	32439113	32439210	*WT1*	e4
chr11	32449492	32449614	*WT1*	e3
chr11	32450033	32450175	*WT1*	e2
chr11	32456236	32456901	*WT1*	e1
chr11	85956262	85956395	*EED*	e1
chr11	85961328	85961500	*EED*	e2
chr11	85963180	85963292	*EED*	e3
chr11	85966254	85966339	*EED*	e4
chr11	85967419	85967564	*EED*	e5
chr11	85968547	85968648	*EED*	e6
chr11	85975204	85975315	*EED*	e7
chr11	85977115	85977268	*EED*	e8
chr11	85979488	85979613	*EED*	e9
chr11	85988012	85988190	*EED*	e10
chr11	85988950	85989043	*EED*	e11
chr11	85989431	85989577	*EED*	e12
chr11	119148866	119149017	*CBL*	e8
chr11	119149210	119149433	*CBL*	e9
chr12	11803052	11803104	*ETV6*	e1
chr12	11905374	11905523	*ETV6*	e2
chr12	11992064	11992248	*ETV6*	e3
chr12	12006351	12006505	*ETV6*	e4
chr12	12022348	12022913	*ETV6*	e5
chr12	12037369	12037531	*ETV6*	e6
chr12	12038850	12038970	*ETV6*	e7
chr12	12043865	12043990	*ETV6*	e8
chr12	25368365	25368504	*KRAS*	e5
chr12	25378538	25378717	*KRAS*	e4
chr12	25380158	25380356	*KRAS*	e3
chr12	25398198	25398328	*KRAS*	e2
chr12	50024381	50024418	*PRPF40B*	e1
chr12	50025174	50025337	*PRPF40B*	e2
chr12	50025633	50025718	*PRPF40B*	e3
chr12	50026369	50026416	*PRPF40B*	e4
chr12	50026628	50026673	*PRPF40B*	e5
chr12	50026787	50026917	*PRPF40B*	e6
chr12	50027200	50027340	*PRPF40B*	e7
chr12	50027410	50027454	*PRPF40B*	e8
chr12	50027659	50027885	*PRPF40B*	e9
chr12	50028105	50028248	*PRPF40B*	e10
chr12	50028311	50028395	*PRPF40B*	e11
chr12	50028872	50029056	*PRPF40B*	e12
chr12	50029138	50029266	*PRPF40B*	e13
chr12	50029616	50029786	*PRPF40B*	e14
chr12	50030489	50030642	*PRPF40B*	e15
chr12	50031243	50031377	*PRPF40B*	e16
chr12	50031506	50031617	*PRPF40B*	e17
chr12	50035681	50035817	*PRPF40B*	e18
chr12	50036008	50036165	*PRPF40B*	e19
chr12	50036353	50036468	*PRPF40B*	e20
chr12	50036700	50036809	*PRPF40B*	e21
chr12	50036996	50037190	*PRPF40B*	e22

Supplemental Table 4.1, continued

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr12	50037464	50037545	*PRPF40B*	e23
chr12	50037634	50037797	*PRPF40B*	e24
chr12	50037873	50037985	*PRPF40B*	e25
chr12	112856906	112856939	*PTPN11*	e1
chr12	112884070	112884212	*PTPN11*	e2
chr12	112888112	112888326	*PTPN11*	e3
chr12	112890989	112891201	*PTPN11*	e4
chr12	112892358	112892494	*PTPN11*	e5
chr12	112893744	112893877	*PTPN11*	e6
chr12	112910738	112910854	*PTPN11*	e7
chr12	112915445	112915544	*PTPN11*	e8
chr12	112915651	112915829	*PTPN11*	e9
chr12	112919868	112920019	*PTPN11*	e10
chr12	112924269	112924443	*PTPN11*	e11
chr12	112926237	112926324	*PTPN11*	e12
chr12	112926818	112926989	*PTPN11*	e13
chr12	112939938	112940070	*PTPN11*	e14
chr12	112942489	112942578	*PTPN11*	e15
chr13	28592594	28592736	*FLT3*	e20
chr13	28608014	28608138	*FLT3*	e15
chr13	28608209	28608361	*FLT3*	e14
chr15	90631809	90631989	*IDH2*	e4
chr17	1554086	1554260	*PRPF8*	e43
chr17	1554392	1554614	*PRPF8*	e42
chr17	1554698	1554857	*PRPF8*	e41
chr17	1554932	1555092	*PRPF8*	e40
chr17	1556826	1556987	*PRPF8*	e39
chr17	1557061	1557320	*PRPF8*	e38
chr17	1558634	1558847	*PRPF8*	e37
chr17	1559676	1559869	*PRPF8*	e36
chr17	1559932	1560065	*PRPF8*	e35
chr17	1561537	1561685	*PRPF8*	e34
chr17	1561810	1562067	*PRPF8*	e33
chr17	1562641	1562852	*PRPF8*	e32
chr17	1563125	1563305	*PRPF8*	e31
chr17	1563716	1563882	*PRPF8*	e30
chr17	1563982	1564131	*PRPF8*	e29
chr17	1564277	1564466	*PRPF8*	e28
chr17	1564555	1564710	*PRPF8*	e27
chr17	1564895	1565094	*PRPF8*	e26
chr17	1565190	1565457	*PRPF8*	e25
chr17	1576365	1576501	*PRPF8*	e24
chr17	1576641	1576871	*PRPF8*	e23
chr17	1577030	1577196	*PRPF8*	e22
chr17	1577726	1577984	*PRPF8*	e21
chr17	1578436	1578643	*PRPF8*	e20
chr17	1578904	1579116	*PRPF8*	e19
chr17	1579212	1579358	*PRPF8*	e18
chr17	1579491	1579674	*PRPF8*	e17
chr17	1579789	1580015	*PRPF8*	e16
chr17	1580260	1580476	*PRPF8*	e15
chr17	1580849	1580998	*PRPF8*	e14
chr17	1581802	1581956	*PRPF8*	e13
chr17	1582046	1582185	*PRPF8*	e12
chr17	1582301	1582510	*PRPF8*	e11

Supplemental Table 4.1, continued

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr17	1582575	1582714	*PRPF8*	e10
chr17	1582893	1583103	*PRPF8*	e9
chr17	1584010	1584135	*PRPF8*	e8
chr17	1584213	1584358	*PRPF8*	e7
chr17	1584762	1584994	*PRPF8*	e6
chr17	1585104	1585342	*PRPF8*	e5
chr17	1585413	1585597	*PRPF8*	e4
chr17	1586817	1587005	*PRPF8*	e3
chr17	1587756	1587875	*PRPF8*	e2
chr17	7572917	7573018	*TP53*	e11
chr17	7573917	7574043	*TP53*	e10
chr17	7576843	7576936	*TP53*	e9
chr17	7577009	7577165	*TP53*	e8
chr17	7577489	7577618	*TP53*	e7
chr17	7578167	7578299	*TP53*	e6
chr17	7578361	7578564	*TP53*	e5
chr17	7579302	7579600	*TP53*	e4
chr17	7579690	7579731	*TP53*	e3
chr17	7579829	7579922	*TP53*	e2
chr17	7748863	7749019	*KDM6B*	e4
chr17	7749180	7749298	*KDM6B*	e5
chr17	7749386	7749625	*KDM6B*	e6
chr17	7749708	7749820	*KDM6B*	e7
chr17	7749887	7750068	*KDM6B*	e8
chr17	7750127	7750344	*KDM6B*	e9
chr17	7750413	7750780	*KDM6B*	e10
chr17	7750854	7753056	*KDM6B*	e11
chr17	7753129	7753275	*KDM6B*	e12
chr17	7753380	7753505	*KDM6B*	e13
chr17	7754329	7754554	*KDM6B*	e14
chr17	7754638	7754722	*KDM6B*	e15
chr17	7754786	7754877	*KDM6B*	e16
chr17	7754956	7755124	*KDM6B*	e17
chr17	7755259	7755393	*KDM6B*	e18
chr17	7755457	7755664	*KDM6B*	e19
chr17	7755803	7755964	*KDM6B*	e20
chr17	7756308	7756454	*KDM6B*	e21
chr17	7756518	7756849	*KDM6B*	e22
chr17	29422318	29422397	*NF1*	e1
chr17	29482991	29483154	*NF1*	e2
chr17	29486018	29486121	*NF1*	e3
chr17	29490194	29490404	*NF1*	e4
chr17	29496899	29497025	*NF1*	e5
chr17	29508430	29508517	*NF1*	e6
chr17	29508718	29508813	*NF1*	e7
chr17	29509516	29509693	*NF1*	e8
chr17	29527430	29527623	*NF1*	e9
chr17	29528045	29528187	*NF1*	e10
chr17	29528419	29528513	*NF1*	e11
chr17	29533248	29533399	*NF1*	e12
chr17	29541459	29541613	*NF1*	e13
chr17	29546013	29546146	*NF1*	e14
chr17	29548858	29548957	*NF1*	e15
chr17	29550452	29550595	*NF1*	e16
chr17	29552103	29552278	*NF1*	e17

Supplemental Table 4.1, continued

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr17	29553443	29553712	*NF1*	e18
chr17	29554226	29554319	*NF1*	e19
chr17	29554531	29554634	*NF1*	e20
chr17	29556033	29556493	*NF1*	e21
chr17	29556843	29557002	*NF1*	e22
chr17	29557268	29557410	*NF1*	e23
chr17	29557850	29557953	*NF1*	e24
chr17	29559081	29559217	*NF1*	e25
chr17	29559708	29559909	*NF1*	e26
chr17	29560010	29560241	*NF1*	e27
chr17	29562619	29562800	*NF1*	e28
chr17	29562926	29563049	*NF1*	e29
chr17	29575992	29576147	*NF1*	e30
chr17	29579946	29580028	*NF1*	e31
chr17	29585352	29585530	*NF1*	e32
chr17	29586040	29586157	*NF1*	e33
chr17	29587377	29587543	*NF1*	e34
chr17	29588719	29588885	*NF1*	e35
chr17	29592237	29592367	*NF1*	e36
chr17	29652828	29653280	*NF1*	e37
chr17	29654507	29654867	*NF1*	e38
chr17	29657304	29657526	*NF1*	e39
chr17	29661846	29662059	*NF1*	e40
chr17	29663341	29663501	*NF1*	e41
chr17	29663643	29663942	*NF1*	e42
chr17	29664376	29664610	*NF1*	e43
chr17	29664827	29664908	*NF1*	e44
chr17	29665033	29665167	*NF1*	e45
chr17	29665712	29665833	*NF1*	e46
chr17	29667513	29667673	*NF1*	e47
chr17	29670017	29670163	*NF1*	e48
chr17	29676128	29676279	*NF1*	e49
chr17	29677191	29677346	*NF1*	e50
chr17	29679265	29679442	*NF1*	e51
chr17	29683468	29683610	*NF1*	e52
chr17	29683968	29684118	*NF1*	e53
chr17	29684277	29684397	*NF1*	e54
chr17	29685488	29685650	*NF1*	e55
chr17	29685977	29686043	*NF1*	e56
chr17	29687495	29687731	*NF1*	e57
chr17	29701021	29701183	*NF1*	e58
chr17	30264256	30264549	*SUZ12*	e1
chr17	30267295	30267361	*SUZ12*	e2
chr17	30267431	30267515	*SUZ12*	e3
chr17	30274626	30274714	*SUZ12*	e4
chr17	30293156	30293225	*SUZ12*	e5
chr17	30300155	30300260	*SUZ12*	e6
chr17	30302491	30302742	*SUZ12*	e7
chr17	30303530	30303643	*SUZ12*	e8
chr17	30310008	30310133	*SUZ12*	e9
chr17	30315329	30315526	*SUZ12*	e10
chr17	30320251	30320362	*SUZ12*	e11
chr17	30320874	30321037	*SUZ12*	e12
chr17	30321573	30321750	*SUZ12*	e13
chr17	30322573	30322791	*SUZ12*	e14

Supplemental Table 4.1, continued

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr17	30323807	30323906	*SUZ12*	e15
chr17	30325667	30326032	*SUZ12*	e16
chr17	74732233	74732556	*SRSF2*	e2
chr17	74732871	74733252	*SRSF2*	e1
chr19	33792234	33793330	*CEBPA*	e1
chr19	56166461	56166529	*U2AF2*	e1
chr19	56170566	56170721	*U2AF2*	e2
chr19	56171533	56171597	*U2AF2*	e3
chr19	56171872	56171995	*U2AF2*	e4
chr19	56172394	56172565	*U2AF2*	e5
chr19	56173858	56173994	*U2AF2*	e6
chr19	56174962	56175120	*U2AF2*	e7
chr19	56179863	56179962	*U2AF2*	e8
chr19	56180026	56180168	*U2AF2*	e9
chr19	56180439	56180557	*U2AF2*	e10
chr19	56180800	56181068	*U2AF2*	e11
chr19	56185290	56185444	*U2AF2*	e12
chr20	30946569	30946645	*ASXL1*	e1
chr20	30954177	30954279	*ASXL1*	e2
chr20	30956808	30956936	*ASXL1*	e4
chr20	31015921	31016061	*ASXL1*	e5
chr20	31016118	31016235	*ASXL1*	e6
chr20	31017131	31017244	*ASXL1*	e7
chr20	31017694	31017866	*ASXL1*	e8
chr20	31019114	31019297	*ASXL1*	e9
chr20	31019376	31019492	*ASXL1*	e10
chr20	31020673	31020798	*ASXL1*	e11
chr20	31021077	31021730	*ASXL1*	e12
chr20	31022225	31025151	*ASXL1*	e13
chr20	42295914	42295953	*MYBL2*	e1
chr20	42302436	42302549	*MYBL2*	e2
chr20	42310414	42310505	*MYBL2*	e3
chr20	42311424	42311536	*MYBL2*	e4
chr20	42315482	42315722	*MYBL2*	e5
chr20	42320787	42320969	*MYBL2*	e6
chr20	42328387	42328694	*MYBL2*	e7
chr20	42331120	42331553	*MYBL2*	e8
chr20	42333849	42334008	*MYBL2*	e9
chr20	42338593	42338712	*MYBL2*	e10
chr20	42340118	42340251	*MYBL2*	e11
chr20	42341632	42341756	*MYBL2*	e12
chr20	42343764	42343933	*MYBL2*	e13
chr20	42344589	42344737	*MYBL2*	e14
chr20	57484395	57484488	*GNAS*	e8
chr20	57484566	57484644	*GNAS*	e9
chr21	36164422	36164917	*RUNX1*	e8
chr21	36171588	36171769	*RUNX1*	e7
chr21	36206697	36206908	*RUNX1*	e6
chr21	36231761	36231885	*RUNX1*	e5
chr21	36252844	36253020	*RUNX1*	e4
chr21	36259130	36259403	*RUNX1*	e3
chr21	36265212	36265270	*RUNX1*	e2
chr21	36421129	36421206	*RUNX1*	e1
chr21	44513202	44513369	*U2AF1*	e8
chr21	44514571	44514683	*U2AF1*	e7

Supplemental Table 4.1, continued

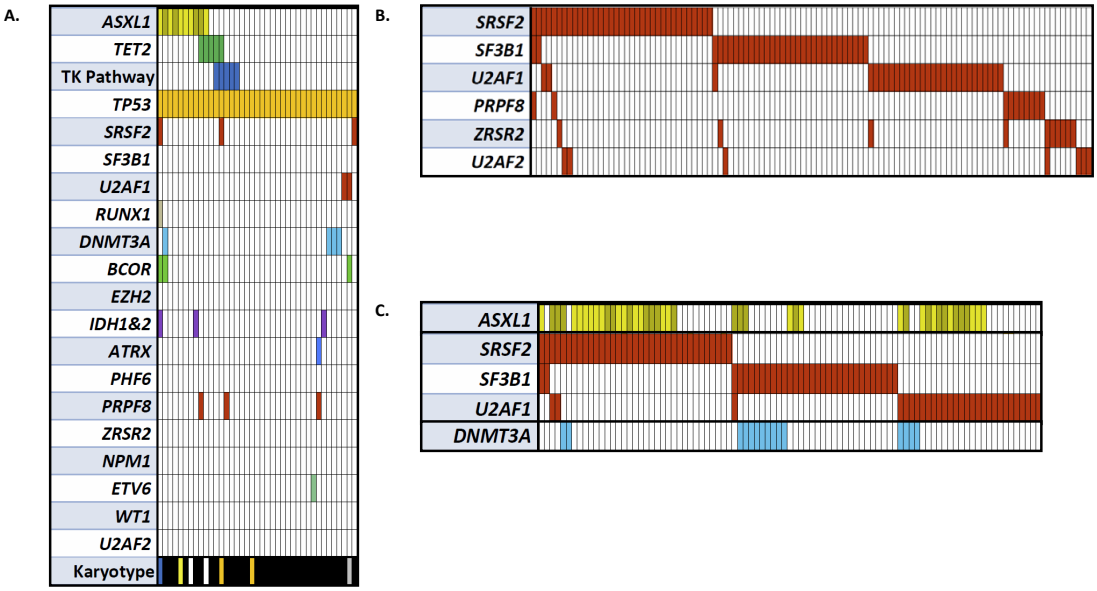
Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chr21	44514755	44514908	*U2AF1*	e6
chr21	44515538	44515656	*U2AF1*	e5
chr21	44515794	44515863	*U2AF1*	e4
chr21	44521466	44521552	*U2AF1*	e3
chr21	44524415	44524522	*U2AF1*	e2
chr21	44527551	44527614	*U2AF1*	e1
chr22	30730573	30730694	*SF3A1*	e16
chr22	30731446	30731537	*SF3A1*	e15
chr22	30731631	30731752	*SF3A1*	e14
chr22	30733005	30733179	*SF3A1*	e13
chr22	30733669	30733896	*SF3A1*	e12
chr22	30734768	30735033	*SF3A1*	e11
chr22	30735109	30735250	*SF3A1*	e10
chr22	30736175	30736380	*SF3A1*	e9
chr22	30736674	30736811	*SF3A1*	e8
chr22	30737671	30737884	*SF3A1*	e7
chr22	30738179	30738349	*SF3A1*	e6
chr22	30738784	30738878	*SF3A1*	e5
chr22	30740912	30741189	*SF3A1*	e4
chr22	30742291	30742518	*SF3A1*	e3
chr22	30748930	30749071	*SF3A1*	e2
chr22	30752709	30752791	*SF3A1*	e1
chrX	39910501	39911653	*BCOR*	e15
chrX	39913139	39913295	*BCOR*	e14
chrX	39913509	39913586	*BCOR*	e13
chrX	39914621	39914766	*BCOR*	e12
chrX	39916408	39916574	*BCOR*	e11
chrX	39921392	39921646	*BCOR*	e10
chrX	39921999	39922324	*BCOR*	e9
chrX	39922861	39923205	*BCOR*	e8
chrX	39923589	39923852	*BCOR*	e7
chrX	39930226	39930412	*BCOR*	e6
chrX	39930890	39930943	*BCOR*	e5
chrX	39931602	39934433	*BCOR*	e4
chrX	39935707	39935785	*BCOR*	e3
chrX	39937097	39937222	*BCOR*	e2
chrX	15808609	15808669	*ZRSR2*	e1
chrX	15809047	15809146	*ZRSR2*	e2
chrX	15817985	15818086	*ZRSR2*	e3
chrX	15821801	15821929	*ZRSR2*	e4
chrX	15822224	15822330	*ZRSR2*	e5
chrX	15826346	15826404	*ZRSR2*	e6
chrX	15827313	15827451	*ZRSR2*	e7
chrX	15833790	15834023	*ZRSR2*	e8
chrX	15836700	15836775	*ZRSR2*	e9
chrX	15838320	15838449	*ZRSR2*	e10
chrX	15840844	15841375	*ZRSR2*	e11
chrX	76763819	76764117	*ATRX*	e35
chrX	76776256	76776404	*ATRX*	e34
chrX	76776871	76776986	*ATRX*	e33
chrX	76777731	76777876	*ATRX*	e32
chrX	76778720	76778889	*ATRX*	e31
chrX	76812912	76813126	*ATRX*	e30
chrX	76814130	76814327	*ATRX*	e29
chrX	76829705	76829833	*ATRX*	e28

Supplemental Table 4.1, continued

Chromosome	Coordinate Start	Coordinate End	Gene Name	Exon #
chrX	76845294	76845420	*ATRX*	e27
chrX	76849156	76849329	*ATRX*	e26
chrX	76854870	76855059	*ATRX*	e25
chrX	76855191	76855299	*ATRX*	e24
chrX	76855893	76856043	*ATRX*	e23
chrX	76872071	76872208	*ATRX*	e22
chrX	76874264	76874459	*ATRX*	e21
chrX	76875853	76876010	*ATRX*	e20
chrX	76888685	76888882	*ATRX*	e19
chrX	76889044	76889210	*ATRX*	e18
chrX	76890075	76890204	*ATRX*	e17
chrX	76891396	76891557	*ATRX*	e16
chrX	76907594	76907853	*ATRX*	e15
chrX	76909578	76909700	*ATRX*	e14
chrX	76912040	76912153	*ATRX*	e13
chrX	76918861	76919057	*ATRX*	e12
chrX	76920124	76920277	*ATRX*	e11
chrX	76931711	76931803	*ATRX*	e10
chrX	76937002	76940095	*ATRX*	e9
chrX	76940421	76940508	*ATRX*	e8
chrX	76944301	76944430	*ATRX*	e7
chrX	76949303	76949436	*ATRX*	e6
chrX	76952055	76952202	*ATRX*	e5
chrX	76953061	76953133	*ATRX*	e4
chrX	76954052	76954127	*ATRX*	e3
chrX	76972598	76972730	*ATRX*	e2
chrX	77041458	77041497	*ATRX*	e1
chrX	133511638	133511795	*PHF6*	e2
chrX	133512025	133512146	*PHF6*	e3
chrX	133527521	133527674	*PHF6*	e4
chrX	133527929	133527992	*PHF6*	e5
chrX	133547511	133547697	*PHF6*	e6
chrX	133547843	133548006	*PHF6*	e7
chrX	133549036	133549160	*PHF6*	e8
chrX	133551189	133551342	*PHF6*	e9
chrX	133559221	133559370	*PHF6*	e10

Supplemental Table 4.2: Overall Response Rates By Mutated Gene and VAF.

Gene	All Mutations			Mutations VAF ≥ 10%		
	N Mutations (%)	WT Response (%)	Mutated Response (%)	N Mutations (%)	WT Response (%)	Mutated Response (%)
ASXL1	99 (46)	57 (50)	43 (43)	91 (43)	62 (51)	38 (42)
TET2	58 (27)	68 (44)	32 (55)	50 (23)	70 (43)	30 (60)
RUNX1	42 (20)	78 (46)	22 (52)	38 (18)	80 (46)	20 (53)
TP53	39 (18)	80 (46)	20 (51)	35 (16)	84 (47)	16 (46)
SRSF2	35 (16)	81 (46)	19 (54)	34 (16)	82 (46)	18 (53)
DNMT3A	34 (16)	85 (47)	15 (44)	30 (14)	87 (48)	13 (43)
SF3B1	32 (15)	87 (48)	13 (41)	31 (15)	88 (48)	12 (39)
U2AF1	29 (14)	87 (47)	13 (45)	27 (13)	88 (47)	12 (44)
NRAS	24 (11)	85 (45)	15 (63)	14 (7)	90 (45)	10 (71)
BCOR	22 (10)	88 (46)	12 (54)	20 (9)	89 (46)	11(55)
NF1	21 (10)	93 (48)	7 (33)	16 (8)	94 (48)	6 (38)
EZH2	21 (10)	90 (47)	10 (48)	20 (9)	91 (47)	9 (45)
IDH2	15 (7)	96 (48)	4 (27)	9 (4)	97 (48)	3 (33)
KDM6B	14 (7)	94 (47)	6 (43)	12 (6)	94 (47)	6 (50)
CBL	14 (7)	98 (49)	2 (14)	10 (5)	98 (48)	2 (20)
ATRX	12 (6)	97 (48)	3 (25)	7 (3)	99 (48)	1 (14)
PHF6	11 (5)	94 (47)	6 (55)	8 (4)	95 (46)	5 (63)
PRPF8	10 (5)	96 (47)	4 (40)	7 (3)	98 (48)	2 (29)
PTPN11	10 (5)	97 (48)	3 (30)	2 (1)	99 (47)	1 (50)
ZRSR2	10 (5)	95 (47)	5 (50)	8 (4)	96 (47)	4 (50)
IDH1	10 (5)	97 (48)	3 (30)	8 (4)	97 (47)	3 (38)
JAK2	10 (5)	96 (47)	4 (40)	8 (4)	97 (47)	3 (38)
KRAS	8 (4)	98 (48)	2 (25)	6 (3)	98 (47)	2 (33)
NPM1	8 (4)	94 (46)	6 (75)	7 (3)	94 (46)	6 (86)
ETV6	8 (4)	95 (46)	5 (63)	7 (3)	96 (47)	4 (57)
WT1	7 (3)	94 (46)	6 (86)	7 (3)	94 (46)	6 (86)
U2AF2	7 (3)	97 (47)	3 (43)	5 (2)	98 (47)	2 (2)
GATA2	6 (3)	96 (46)	4 (67)	6 (3)	96 (46)	4 (67)
MYBL2	5 (2)	97 (47)	3 (60)	5 (2)	97 (47)	3 (60)
SUZ12	5 (2)	98 (47)	2 (40)	5 (2)	98 (47)	2 (40)
CEBPA	4 (2)	99 (47)	1 (25)	2 (1)	100 (47)	0 (0)
EED	3 (1)	100 (48)	0 (0)	3 (1)	100 (48)	0 (0)
PRPF40B	3 (1)	99 (47)	1 (33)	2 (1)	99 (47)	1 (50)
LUC7L2	2 (1)	98 (46)	2 (100)	2 (1)	98 (46)	2 (100)
SF3A1	2 (1)	98 (46)	2 (100)	2 (1)	98 (46)	2 (100)
MPL	2 (1)	100 (47)	0 (0)	0 (0)	100 (48)	0 (0)
CBLB	2 (1)	98 (46)	2 (100)	1 (<1)	99 (47)	1 (100)
BRAF	1 (<1)	99 (47)	1 (100)	0 (0)	100 (48)	0 (0)
GNAS	1 (<1)	100 (47)	0 (0)	1 (<1)	100 (47)	0 (0)
KIT	1 (<1)	99 (47)	1 (100)	1 (<1)	99 (47)	1 (100)



D.

	CMML (n)	CMML Frequency	non-CMML Frequency
ASXL1	16	76%	43%
TET2	8	38%	26%
TK Pathway	10	48%	25%
TP53	3	14%	19%
SRSF2	7	33%	15%
SF3B1	1	5%	16%
U2AF1	3	14%	14%
RUNX1	6	29%	19%
DNMT3A	2	10%	17%
BCOR	2	10%	10%
EZH2	6	29%	8%
IDH1&2	3	14%	11%
ATRX	1	5%	6%
PHF6	2	10%	5%
PRPF8	1	5%	5%
ZRSR2	1	5%	5%
NPM1	0	0%	4%
ETV6	2	10%	3%
WT1	1	5%	3%
U2AF2	1	5%	3%

E.

	RARS (n)	RARS Frequency	non-RARS Frequency
ASXL1	5	21%	50%
TET2	9	38%	26%
TK Pathway	5	21%	28%
TP53	1	4%	20%
SRSF2	1	4%	18%
SF3B1	16	67%	8%
U2AF1	4	17%	13%
RUNX1	6	25%	19%
DNMT3A	5	21%	15%
BCOR	7	29%	8%
EZH2	3	13%	10%
IDH1&2	1	4%	12%
ATRX	3	13%	5%
PHF6	1	4%	5%
PRPF8	1	4%	5%
ZRSR2	0	0%	5%
NPM1	0	0%	4%
ETV6	1	4%	4%
WT1	1	4%	3%
U2AF2	0	0%	4%

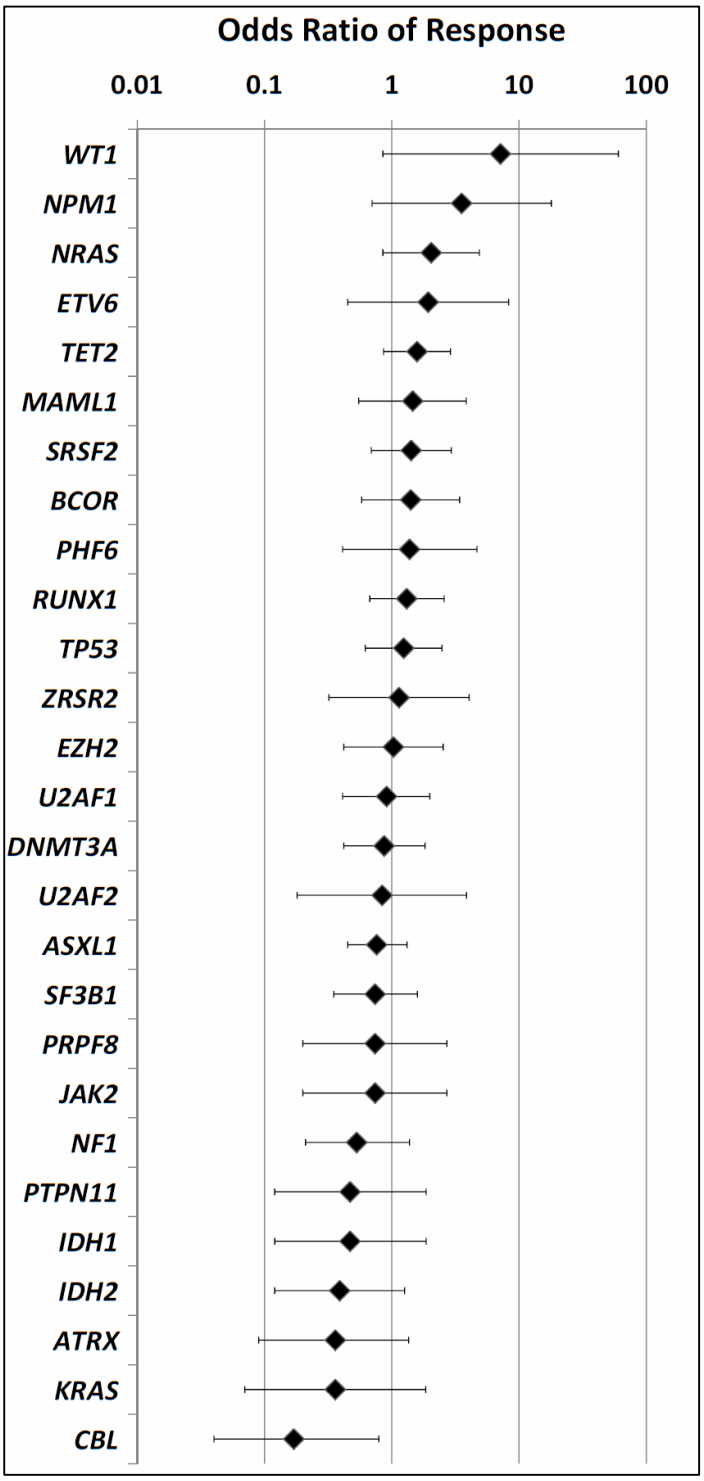
Supplemental Figure 4.1: Patterns of mutations consistent with results from other published MDS cohorts.

Each column represents a patient sample. Each colored bar indicated a mutation of the gene or genes described in that row to the left. A) Patients with TP53 mutations have a lower frequency of mutations in other genes compared to TP53 wildtype patients and are strongly associated with complex karyotypes (black bars shown in karyotype row). B) Mutations in splicing factor genes are largely mutually exclusive. (Continued next page)

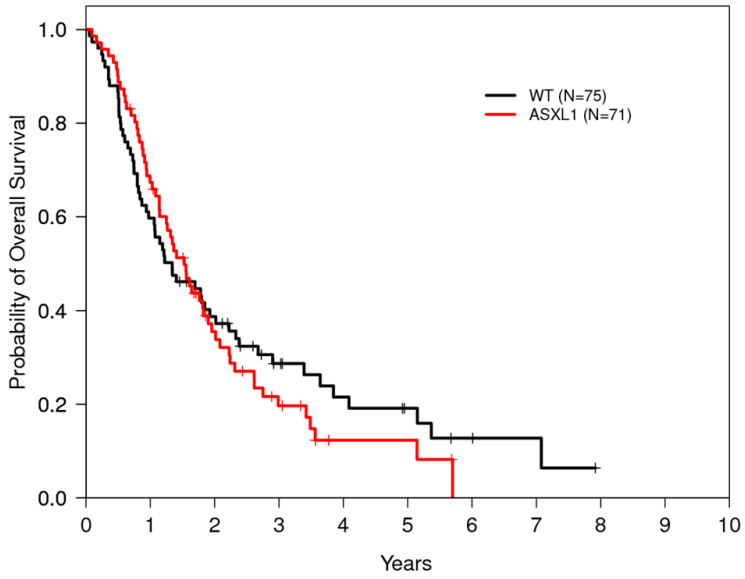
Supplemental Figure 4.1, continued:

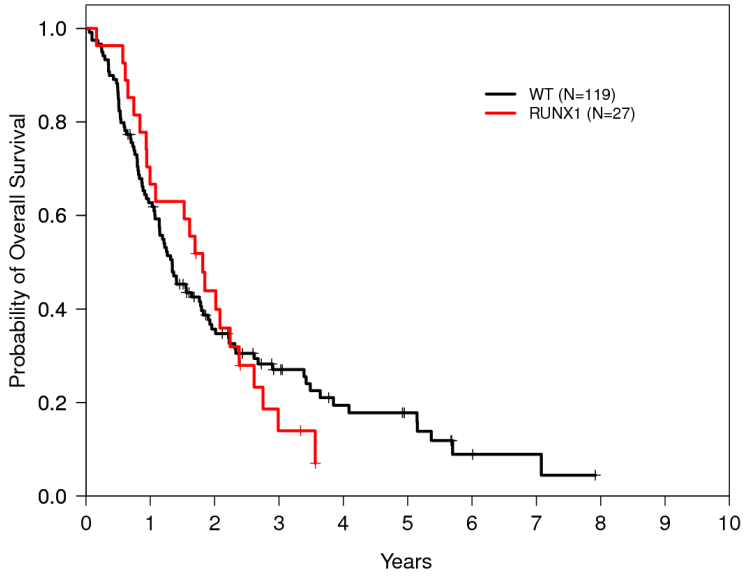
C) Patients with SF3B1 mutations have lower rates of concurrent ASXL1 mutations and higher rates of concurrent DNMT3A mutations. D) Mutation profile in CMML patients shows higher incidence of mutations in ASXL1, TK Pathway genes, SRSF2, and EZH2 compared with non-CMML patients. E) Mutation profile in RARS patients shows high rate of SF3B1 mutations and lower rate of ASXL1, TP53, and SRSF2 mutations.

TK Pathway refers to mutations in any of the following: NRAS, KRAS, BRAF, CBL, JAK2, PTPN11, MPL, KIT, or CBLB. Karyotype colors: Black – Complex, Blue – trisomy 8, Yellow – del(20q), White – Normal, Red – del(5q), Green - -7/del(7q), Gray – Other, Orange – Unknown

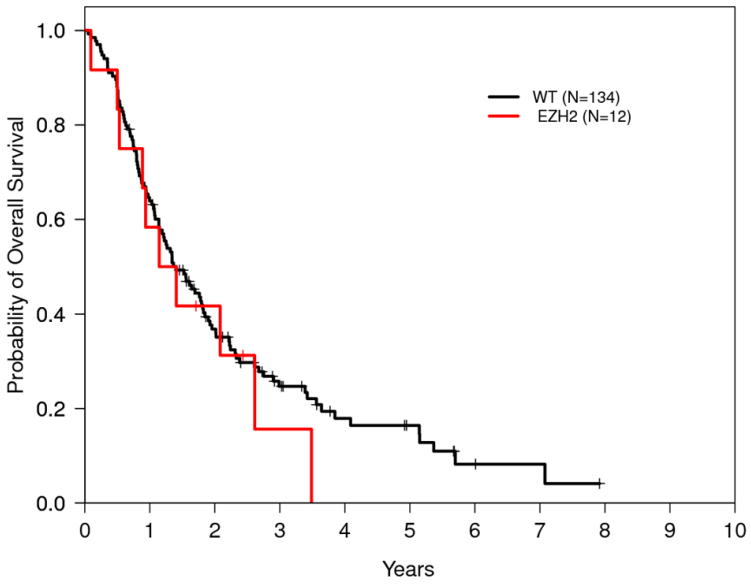


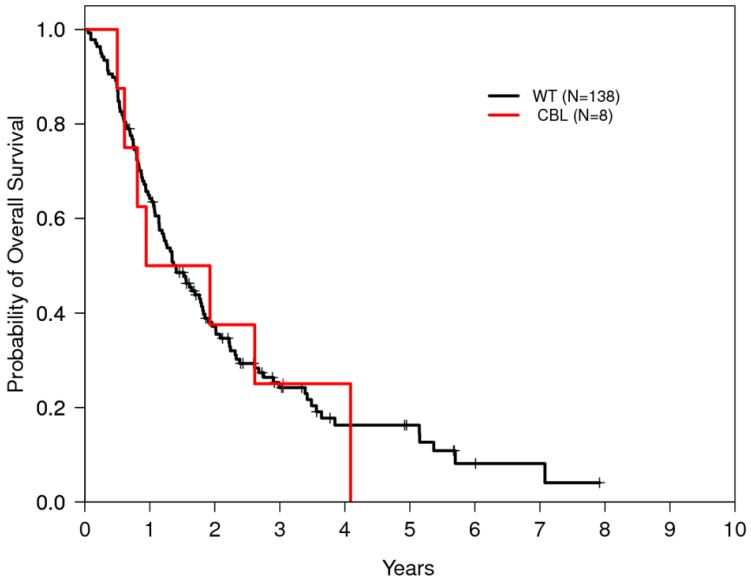
Supplemental Figure 4.2: Univariate odds ratio of response by mutated gene regardless of variant allele fraction is shown with 95% confidence intervals indicated.



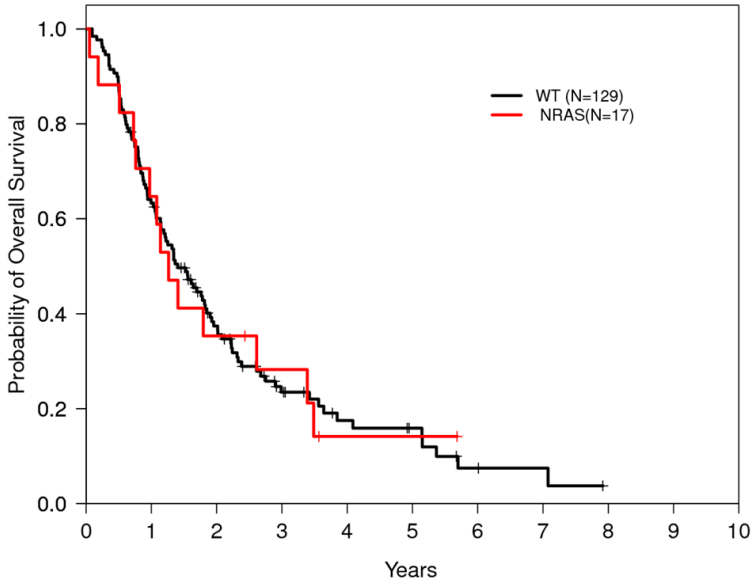


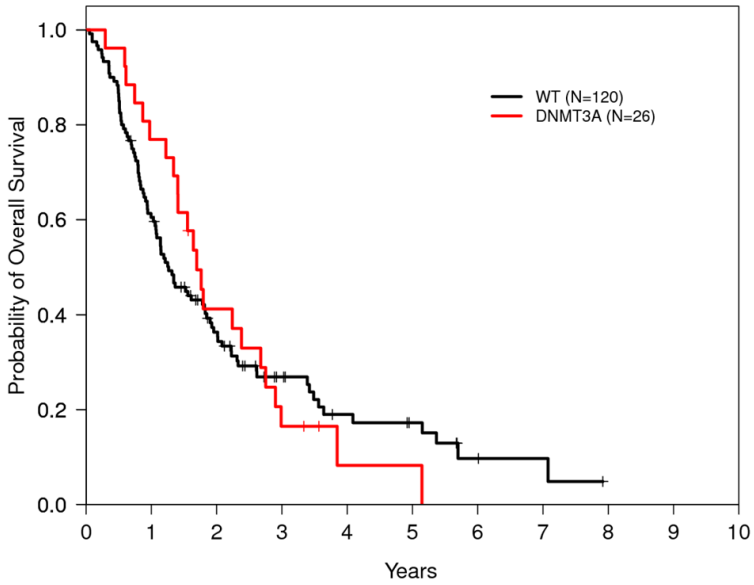
Supplemental Figure 4.3: Kaplan-Meier curves for overall survival in patients with and without mutations in select genes shown to have association with overall survival in other cohorts (continued next page).



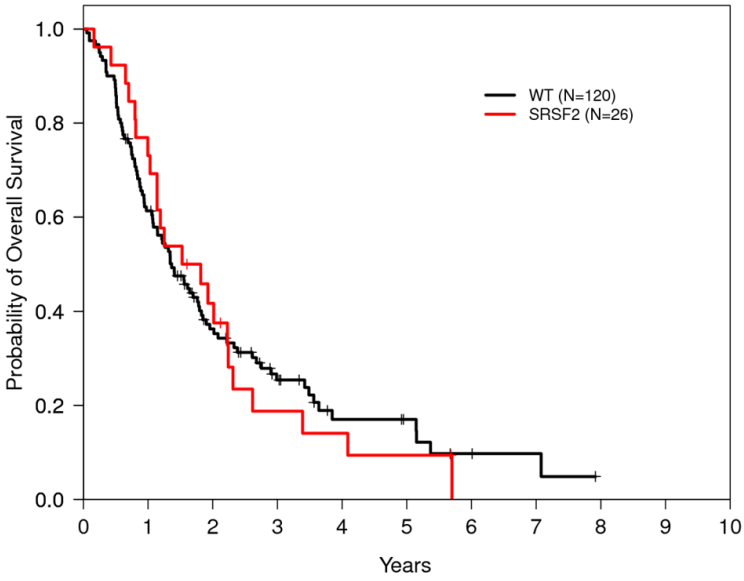


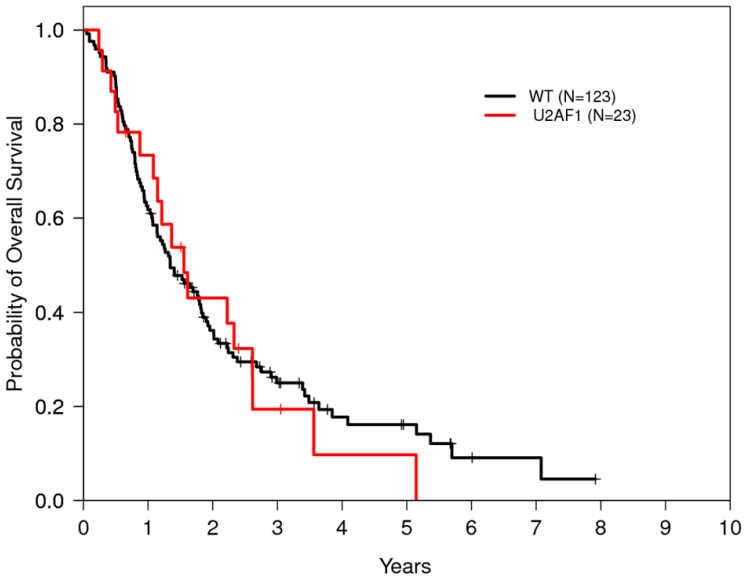
Supplemental Figure 4.3, continued



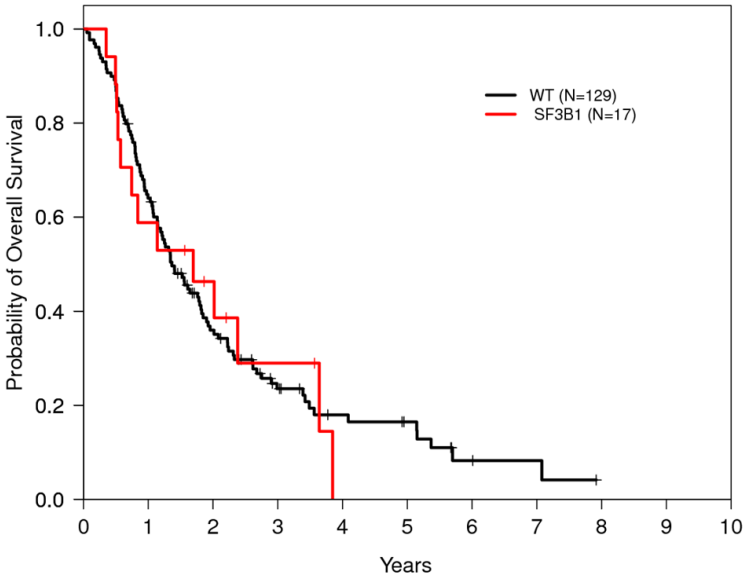


Supplemental Figure 4.3, continued





Supplemental Figure 4.3, continued



Supplemental Figure 4.3, continued